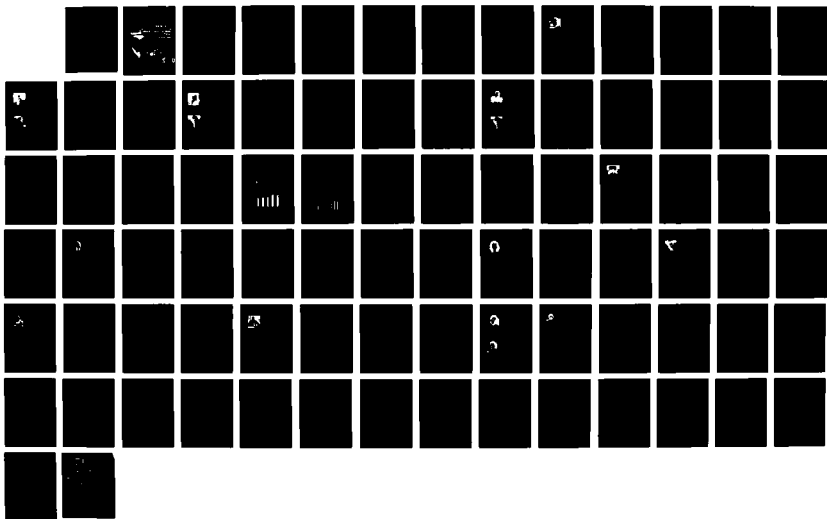


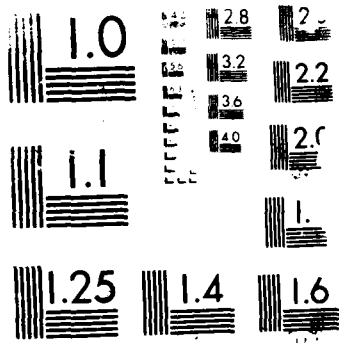
AD-A207 553

INDEPENDENT RESEARCH AND INDEPENDENT EXPLORATORY
DEVELOPMENT PROGRAMS: FY. (U) NAVY PERSONNEL RESEARCH
AND DEVELOPMENT CENTER SAN DIEGO CA.
H E MONTAGUE ET AL. MAR 89 NPRDC-AP-89-7 F/G 5/8 NL

1/1

UNCLASSIFIED





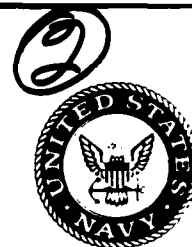
DTIC FILE COPY

Navy Personnel Research and Development Center

San Diego, CA 92152-6800

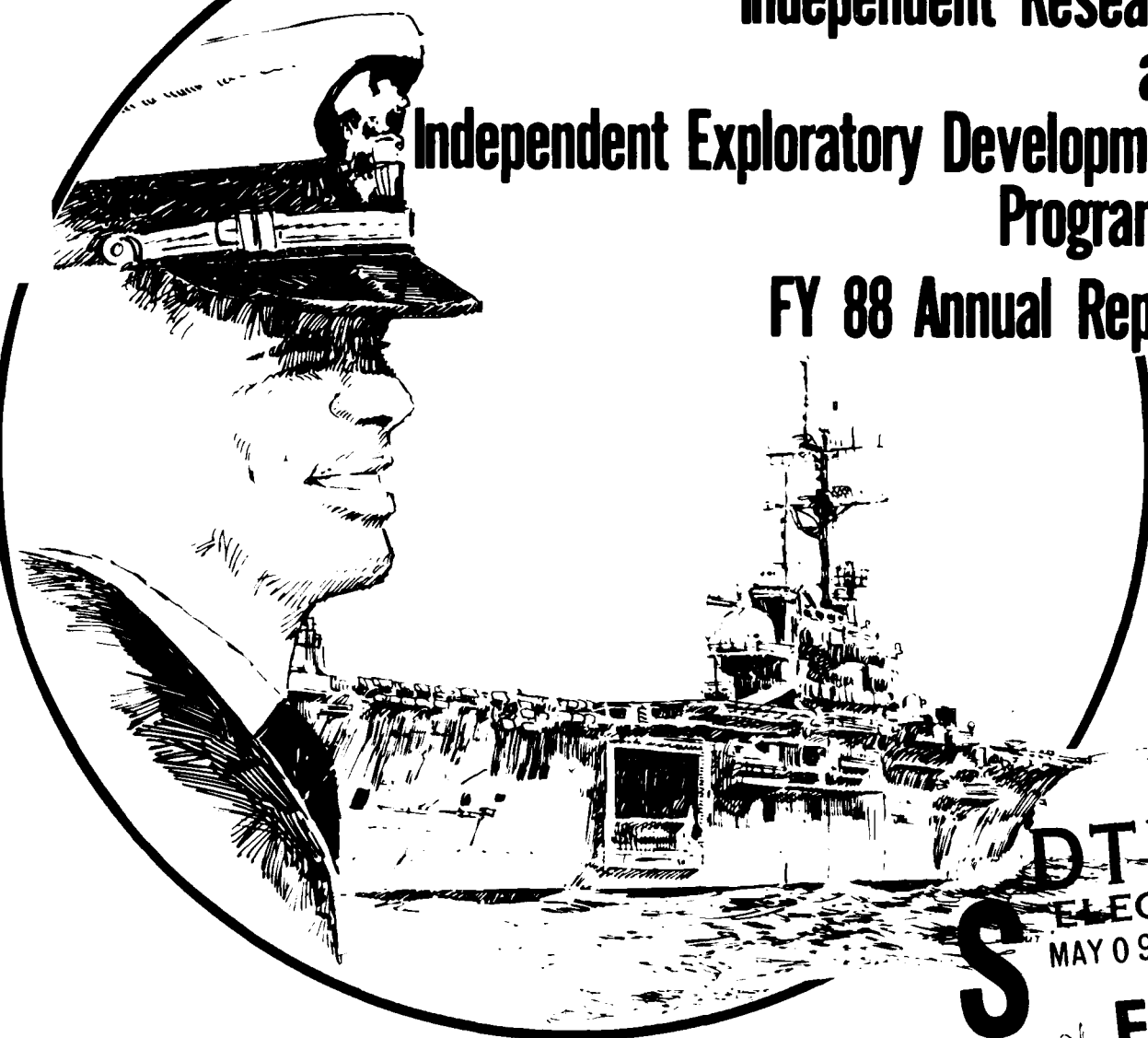
AP 89-7

March 1989



AD-A207 553

Independent Research and Independent Exploratory Development Programs: FY 88 Annual Report



DTIC
ELECTE
MAY 09 1989
S E D

89 00 020

Administrative Information

This report documents the activities and accomplishments of the Independent Research (IR) and Independent Exploratory Development (IED) programs at the Navy Personnel Research and Development Center for FY88. In addition to the technical presentations, program administrative information is provided. For further information, contact the IR/IED Program Coordinator, Dr. William E. Montague, Autovon or (619) 553-7849, or any of the Principal Investigators.

**Independent Research
and
Independent Exploratory Development
Programs:
FY88 Annual Report**

*William E. Montague
Carmen C. Scheifers
(Editors)*

*Reviewed by
J. S. McMichael
Technical Director*

*Approved and Released by
B. E. Bacon
Captain, U.S. Navy
Commanding Officer*

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Approved for public release; distribution is unlimited.



**Navy Personnel Research and Development Center
San Diego, CA 92152-6800**

CONTENTS

Independent Research Program

Page

1	Introduction	
5	Stabilization of Performance on a Computer-based Simulation of a Complex Cognitive Task	Pat-Anthony Federico
10	An Analysis of Tutoring in Technical Training in the Classroom/Laboratory and On-The-Job	John A. Ellis William E. Montague
13	Brain Wave Correlates of Memory Performance	Diane Williams Gregory W. Lewis
18	Brain Mechanisms for Human Color Vision; Implications for Display Systems	Leonard J. Trejo Gregory W. Lewis
34	Experienced-based Career Development	Robert Morrison
39	How to Elicit Knowledge from Experts	Donald Bamber

Independent Exploratory Development Program

46	Reading Comprehension Strategies	Meryl Sue Baker
49	Optimal Control Theory for a System of Quasi-Linear Difference Equations	Iosif Krass
52	Loss Forecasting With Empirical Bayes Estimators	James P. Boyle
56	Models for Calibrating Multiple-choice Items	J. Bradford Sympton
61	Group Size and Member Approval of Reward Plans in a Gain Sharing System; Effects on Individual Performance	Delbert M. Nebeker Paul H. DeYoung B. Charles Tatum

A-0	Appendix A--Project Transitions
B-0	Appendix B--Presentations and Publications
C-0	Appendix C--Honors and Awards
	Distribution List

Introduction

New and innovative ideas proposed by research scientists at the Navy Personnel Research and Development Center (NPRDC) are encouraged by the Technical Director, Dr. James McMichael, to promote scientific and technological growth in the organization and the development of knowledge of interest to the Navy. Support is provided by discretionary funding furnished by the Independent Research (IR) and Independent Exploratory Development (IED) programs of the Office of Naval Research and the Office of Naval Technology. These programs support initial research and development of interest to the Navy with emphasis on the NPRDC mission areas of the acquisition, training, and effective utilization of personnel.

Funds are provided to the Technical Directors of Navy Laboratories to support innovative, promising research and development outside the procedures required under normal funding authorization. The funds are to encourage creative efforts important to mission accomplishment. They enable promising researchers to spend a portion of their time on examining the feasibility of self-generated new ideas and scientific advances. They can provide important and rapid test of promising new technology, and can help fill gaps in the research and development program. This may involve preliminary work on speculative solutions too risky to be funded from existing programs.

The funds also serve as means to maintain and increase the necessary technology base skill levels and build in-house expertise in areas likely to become important in the future. These programs contribute to the scientific base for future improvements in the manpower, personnel, and training systems technology, and provide coupling to university and industrial research communities.

The FY88 IR/IED programs began with a call for proposals in May 1987, which resulted in 21 submissions. Technical reviews were provided by supervisors and scientific consultants and six IR and four IED projects were funded. This report documents the results and accomplishments of these projects. Dr. W. E. Montague, Code 15A, administers the IR and IED programs, coordinating project selection, reporting, and review to assure an innovative and productive program of science and technology.

Tables 1 and 2 provide information on the projects active during FY88 and list those supported in FY89. The subsequent pages contain short reports of research progress during FY88 written by the principle investigators of each project. Appendix A lists the IR and IED projects that may have transitioned into other projects or into use by the Navy during the year. Appendix B presents the presentations and publications and Appendix C the awards.

Table 1
Independent Research
Work Units for FY88 and FY89
(PE 0601152N)

Work Unit	Title	Principal Investigator	Internal Code	Telephone (619) 553- or A/V 533	FY Funding (\$K)	
					88	89 ^a
R000-N0-000-01	Brain mechanisms for human color vision: Implications for display systems	Trejo/Lewis	15	37981/ 37942	67	0
R000-N0-000-02	How to elicit knowledge from experts ^b	Bamber	16	39212	40	--
R000-N0-000-03	Stabilization of performance on a computer-based simulation of a complex cognitive task ^c	Federico	14	37777	75	25
R000-N0-000-04	Experienced-based career development	Morrison	12	39256	45	45
R000-N0-000-05	Event-related potential correlates of memory performance	Williams/ Lewis	15	37925/ 37942	50	20 ^d
R000-N0-000-06	An analysis of tutoring in technical training in the classroom/ laboratory and on-the-job	Ellis/ Montague	14	39273/ 37849	25	60
R000-N0-000-07	Brain activity during visual recognition: Implications for Navy training	Lewis	14	37988	0	60
R000-N0-000-08	Using diagrams for learning procedural tasks	Vogt	15	37788	0	40
R000-N0-000-09	Application of machine intelligence	Sorensen	15	37782	0	45
R000-N0-000-XX	Undistributed as of January 1989		14	37849	8	55
					<u>310</u>	<u>350^a</u>

^aDecember 1988, 50 percent of funds received and each project funded at half amount shown

^bTransferred to Naval Ocean Systems Center

^cTransitioned

^dAdditional support obtained from ONR

Table 2
Independent Exploratory Development
Work Units for FY88 and FY89
(PE 0602936N)

Work Unit	Title	Principal Investigator	Internal Code	Telephone (619) 553- or A/V 533	FY Funding (\$K)	
					88	89
RV36-I27-01	Optimal control theory for a system of quasi-linear difference equations ^a	Krass	11	37962	50	50
RV36-I27-02	Reading comprehension strategies ^{a, b}	Baker	15	39305	30	0
RV36-I27-03	Models for calibrating multiple-choice items ^b	Sympson	13	37610	46	0
RV36-I27-04	Group size and member approval of reward plans in a gain-sharing system: Effects on individual performance	Nebeker/DeYoung/Tatum	16	37749/ 37943/ 37758	60	0
RV36-I27-05	Loss forecasting with empirical Bayes estimators ^c	Boyle	11	38025	40	50
RV36-I27-06	Military recruit quality and the minimum wage	Nakada	15	39268	0	60
RV36-I27-07	Using a neural net approach in manpower forecasting	Huntley	11	37923	0	35
RV36-I27-XX	Undistributed as of January 1989		14	37849	4	35
					<u>226</u>	<u>100</u>

^aTransitioned^bResearch completed^cWill transition

Biography



PAT-ANTHONY FEDERICO is a Senior Research Psychologist in the Training Technology Department. He earned his B.A. *cum laude* from the University of St. Thomas in 1965 with a double major in mathematics and philosophy and a minor in physics. He was awarded his Ph.D. in 1969 from Tulane University in general experimental psychology. He has research interests in individual differences in cognitive processing, learning, and performance; and computer-based instruction and performance assessment. He was elected and

served as Executive Director, President, and Secretary-Treasurer of the Human Factors Society, San Diego Chapter. He is also a member of the Cognitive Science Society, Psychonomic Society, American Educational Research Association, and American Psychological Association. He is a member of the editorial advisory review board for the *Journal of Educational Psychology*, and an ad hoc reviewer for *Human Factors and Memory and Cognition*. He is a peer reviewer and advisor for the Office of the Assistant Secretary for Educational Research and Improvement, United States Department of Education. He has authored or edited over 80 scientific contributions including books, chapters, journal articles, professional papers, and technical reports.

STABILIZATION OF PERFORMANCE ON A COMPUTER-BASED SIMULATION OF A COMPLEX COGNITIVE TASK

Pat-Anthony Federico

The purpose of this research is to study the processes intrinsic to the stabilization of performance on a complex cognitive task (conducting an outer air battle). Subjects interact with an animated, computer-based graphic simulation. They allocate, deploy, and manage tactical assets in a very large number of scenarios to defend carrier-based task forces against hostile, missile-launching bombers. Concurrent and retrospective verbal protocols are obtained from the subjects regarding their battle management. Performance during each scenario is automatically assessed by the computer system against 16 multivariate measures. Cognitive and statistical analyses will be conducted to study the processes of acquiring skill and reaching stabilization of performance on this complicated mental task. Contributions to methodology and theory culminating from this research will result in improved operationally oriented performance assessment.

Background

Individuals vary in their rates and manners of skill acquisition, especially in the beginning of practice, and they reach terminal performance plateaus differentially. Early performance requires high conscious control (i.e., it is slow, sequential, effortful, limited, and directed), whereas late performance tends to be automatic (i.e., it is fast, parallel, effortless, and less limited by attentional focus). Practice during the early stages results in dramatic changes in behavior (e.g., decreasing performance variability, minimizing response time). With practice, rate of improvement diminishes and becomes more uniform across individuals

(i.e., performance stabilizes). For some tasks, performance does not seem to get any better or worse, and curves that reflect the rate of skill acquisition of individuals appear to be parallel (Ackerman & Schneider, 1984; Jones, 1984; Schneider, 1984). Individual variability among learners affects modes and speed of skill acquisition: Distinct experiences, cognitive models, aptitudes, and motivations can influence early and late performance differentially.

Much of the earlier research on which the above statements are based was done with psychomotor tasks. A lot less is known about complex tasks, which are primarily cognitive in nature.

Problem

Because many factors affect the nature and time course of acquisition, beginning performance on complicated tasks is usually not a good estimate of terminal performance. Since usually and initially intricate performance does not stabilize, it may reflect distinct facets of skill on different attempts to perform as indicated above. In other words, estimates of performance are likely to measure different things on different trials for different people. Trying to separate accurately better and poorer performing people, or to determine consistently whether a trainee has mastered a needed skill become difficult. This potential lack of reliability impacts upon the predictive power of computer-based simulations for assessing operationally oriented skills. Therefore, it affects the validity of computer simulations for job-sample-performance testing in functional contexts.

Technological Objective

The technological objective of this proposed research is to conduct cognitive and statistical analyses as well as theoretical modeling to study the process of skill acquisition resulting in the stabilization of performance on a computer-based simulation of a complex cognitive task.

General Approach

Target Task

The target task of this proposed research consists of tactically allocating, deploying, and managing fighter and supporting aircraft to defend an aircraft carrier and its escorting ships against threatening Soviet naval air bombers. This task demands considerable practice

before it can be executed with a sufficiently high level of skill and becomes automatic. For the purposes of this research, this task is considered as a test of individual differences in complex mental performance. In the execution of this task the transition from controlled to automatic performance is important. This implies that what is crucial is not early but late performance (i.e., how well individuals do after extended practice). The administration of numerous trials on this task, together with cognitive and statistical analyses, make it possible to note when and how stabilization of performance is achieved (i.e., when the research subjects no longer show any tendency to improve or worsen with practice).

Computer-based Simulation

Software tools were developed for constructing computer-based animated graphic simulations of the actual radar coverage of F-14 and F/A-18 fighters, KA-6 tankers, and E2-C early warning aircraft as well as fuel flow of these planes. Included is the probability of kill for Phoenix, Sparrow, and Sidewinder missiles that the different fighters carry. It is possible to generate an infinite number of raids from Soviet naval air bombers with antiship missiles (ASMs) in different warfare theaters and various carrier loadouts in terms of numbers of each type of fighter and missile on board enable the creation of an infinite set or universe of tactical scenarios. These tools are used to assess how well individuals manage outer air battles to defend carrier-based naval task forces.

Subjects

The research subjects, approximately six F-14 pilots and radar intercept

officers at Naval Air Station (NAS) Miramar and/or instructors and students from the Tactical Action Officer, Tactical Warfare Overview, and/or Staff Tactical Watch Officer Courses from the Fleet Combat Training Center Pacific, will be required to allocate, deploy, and manage fighter and supporting aircraft in order to knock down various numbers and mixes of hostile bombers before they reach their respective ASM launch points. Each computer-based scenario will be run in compressed or accelerated time; each threat scenario is considered as a performance test item.

Performance Criteria

A subject's tactical performance during simulated air battles is assessed according to 16 multivariate criteria. Some of these are as follows: the percentage of incoming threat aircraft which were detected by F-14, F/A-18, and E2-C radar systems, the percentage of bombers that fighters placed in missile launch acceptability regions (LARs), the percentage of hostile aircraft shot down or probable kills, the average range from the defended task force at which threat aircraft were knocked down, the percentage of hostile platforms knocked down before ASMs were launched, etc.

Procedure

Subjects are run on the computer-based scenarios of these symbolically displayed air battles between Soviet bombers and U.S. carrier-based aircraft. How well each allocates, deploys, and manages fighters and other supporting aircraft during the simulated battle is assessed according to the performance criteria mentioned above. The possible number of incoming raids or specific

threat scenarios form a practically infinite universe. Consequently, the set of simulated tactical scenarios is considered as an operationally oriented, domain-referenced, job-sample, performance test. With each scenario as an assessment trial, subjects are administered 200 trials divided into 20 blocks.

Cognitive Analysis

During the first trial of every block, verbal protocols are obtained from the subjects as they are actually conducting the simulated air battles. The analyses of these concurrent verbalizations, as well as retrospective reports, disclose the information needed by the subjects while they perform this complex task. Comparisons of the thinking-aloud protocols and retrospective reports on the first trial of every block reveal the variability in cognitive processing within as well as between subjects as they acquire skill (i.e., progress from controlled to more automatic performance of the task).

Analysis of protocols obtained early and late during practice on the task indicate how subjects' cognitive processes and structures change as their performances tend to stabilize. These reflect the cognitive correlates of the acquisition of stable task performance. Together with a thorough componential analysis, the information obtained from the protocol analysis will be used to construct a model for performing this complex task. This model will be used to create a theoretical framework as well as build the basis of an expert system for a computer-based "intelligent tactician" that will monitor, diagnose, and assess the conduct of simulated air battles to defend carrier task forces.

Statistical Analyses

Combining statistical procedures with protocol analyses and conceptual modeling provides an integrated account of the cognition accompanying the acquisition of complex task performance. Together with cognitive analysis and theory, statistical techniques (e.g., a test for the homogeneity of k regression lines) can be used to uncover the mental processes and structures underlying the acquisition of stabilization.

Potential Products/Transition

The potential products of this research are contributions to a knowledge base and much needed theoretical framework. The contributions to methodology and theory culminating from this research can be extended or transitioned to the exploratory development of "intelligent or expert" computer-based simulation systems to measure complex cognitive performance in functional contexts. Then, the predictive power of this type of performance assessment can be determined. Likewise, this follow-on work itself can be transitioned to advanced development of an intelligent computer-based simulation system to support job-sample performance assessment of intricate cognitive tasks. This advanced system would allow accessing of developed methodologies,

theoretical orientations, mental models, as well as generic software tools to implement prescriptive procedures to aid in the production of performance tests for complex cognitive tasks.

Progress

Cognitive and statistical performance data have been collected. All verbal protocols have been transcribed. Currently, the remaining transcribed protocols are being coded and inter-rater reliabilities are being computed on a sample of the coded protocols.

References

- Dillon, R. F., & Federico, P-A. (1986). *Innovations in training and testing*. Symposium held at the meeting of the American Educational Research Association, San Francisco, CA.
- Federico, P-A. (1987). Cerebral, cognitive, and conative processes. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction: Volume 3, Conative and affective process analyses*. Hillsdale, NJ: Lawrence Erlbaum Associates.

See page B-3 for FY88 publication.

Biography



JOHN A. ELLIS is a Research Psychologist in the Training Technology Department. He has been involved in research and developments programs dealing with quality control of instructional development, criterion reference testing, cognition and instruction, computer-based training, and instructional systems design. He received his doctorate in Psychology from the University of Illinois, Champaign-Urbana in 1976. He has over 100 publications and professional publications. His current research interests are in traditional classroom training, expert systems for instructional development, and techniques for enhancing the retention of procedural tasks. Dr. Ellis is a member of the American Educational Research Association and the Steering Committee of the Military Testing Association. He is currently a consulting editor for the *Journal of Educational Psychology*, and is a peer review advisor for the Office for Educational Research and Improvement of the Department of Education.



WILLIAM E. MONTAGUE is a Senior Scientist in the Training Technology Department. His research specialty is cognition and learning. For several years he directed projects developing improvements of instructional design methods and using computers for training. Trained as an experimental psychologist at the University of Virginia, he did research in human factors for the Navy Electronics Laboratory, taught Psychology and Educational Psychology at the University of Illinois, and moved to NPRDC in 1972 as a project leader. He is an active member of several professional organizations including: *American Educational Research Association*, *Cognitive Science Society*, *Psychonomic Society*, *American Psychological Association*, *Human Factors Society*, and *Military Testing Association*. He has authored or co-authored over 100 professional and technical papers, and has co-edited three books concerned with instructional psychology. He is currently a consulting editor for the *Journal of Educational Psychology*, the *Journal of Applied Psychology*, the *Human Factors Journal*, and is a peer review advisor for the Office for Educational Research and Improvement of the Department of Education.

AN ANALYSIS OF TUTORING IN TECHNICAL TRAINING IN THE CLASSROOM/LABORATORY AND ON-THE-JOB

John A. Ellis
William E. Montague

In addition to the more than 7000 formal courses taught in Navy schools, there is a considerable amount of training conducted on-the-job in ship and shore based commands. Although the Navy has courses and programs that prepare petty officers to be leaders (e.g., LMET), there is no formal training on how to be on-the-job trainers (i.e., tutors). The goal of this project is to do the basic research required to provide information for designing and developing a formal program for teaching senior Navy petty officers to be effective on-the-job trainers/tutors. The work will proceed in four phases (1) analysis of tutoring, (2) developing a data collection methodology, (3) collecting baseline data on lab instructors in Navy schools (this phase will be done in conjunction with phase 2), and (4) data collection aboard ship.

Background and Problem

In addition to the more than 7000 formal courses taught in Navy schools, there is a considerable amount of training conducted on-the-job in ship and shore based commands. In peace time, the Navy is heavily involved in training. This is especially true for new job incumbents and for those in jobs that change frequently or are difficult to master (i.e., the task are complex, there are infrequent opportunities for practice, etc.) Much of this training occurs informally in one-on-one or one-on-two or three situations, with a senior petty officer (e.g., E-6, E-7) working with/teaching seaman and seaman apprentice personnel on/about shipboard tasks. These senior petty officers are in effect

tutors and are responsible for bringing "A" school (and non "A" school) graduates from a novice status to a journeyman. This involves preparing them to take and pass advancement exams, meet PQS and practical factor requirements, and perform their jobs. Although the Navy has courses and programs that prepare petty officers to be leaders (e.g., LMET), there is no formal training on how to be on-the-job trainers (i.e., tutors).

Objective

The objective of this project is to do the basic research required to provide information for designing and developing a formal program for teaching senior Navy petty officers to be effective on-the-job trainers/tutors.

Approach

The project consists of four phases (1) analysis of tutoring, (2) developing a data collection methodology, (3) collecting baseline data on lab instructors in Navy schools (this phase will be done in conjunction with phase 2), and (4) data collection aboard ship.

Phase 1 involves an analysis of tutoring to determine the factors involved in tutoring and the characteristics of good tutors. Several researchers are currently investigating these issues (e.g., Fox 1988, Gordon 1988) with tutors in college subjects. Phase 1 extends this work to technical training.

In Phase 2, a data collection methodology will be developed for assessing tutorial skills and knowledge. It is anticipated that this methodology will involve ethnographically oriented video taping and field observations, as well as paper-and-pencil surveys and interviews. The methodology would be developed in the laboratory section of a Navy course.

During Phase 3, data will be collected on lab instructors as they work with students. As these instructors are trained to work with students (although not necessarily to be tutors), this data would serve as baseline data for the shipboard observations (Phase 4).

In Phase 4, data will be collected on petty officers aboard ship as they work with junior enlisted personnel. This data will be compared to the lab instructors to determine how effective ship board and school personnel are in

tutoring. The results will be used to make recommendations for a formal training program in tutoring for senior petty officers and for modifications in instructor training to enhance tutoring skills.

Progress

This project began late in the fourth quarter of FY88. Therefore, all that has been accomplished is preliminary work on Phase 1.

Plans

Phases 1, 2, and 3 could be completed in FY89 and Phase 4 could start in FY89 and be completed in FY90.

References

- Fox, B. A. (1988). *Cognitive and interactional aspects of correction in tutoring* (TR 88-2). Boulder: University of Colorado, Institute of Cognitive Science.
- Fox, B. A. (1988). *Interaction as a diagnostic resource in tutoring* (TR 88-3). Boulder: University of Colorado, Institute of Cognitive Science.
- Fox, B. A. (1988). *Repair as a factor in interface design* (TR 88-4). Boulder: University of Colorado, Institute of Cognitive Science.
- Gordon, E. E. (1988). *Description of a field tutoring project and IIP curriculum*. Paper presented at the 1988 annual meeting of the American Educational Research Association, New Orleans, LA.

Biography



DIANE WILLIAMS was born and raised in Tacoma, Washington. She did undergraduate work at the University of Washington, obtaining a B.S. in Mathematics and a B.S. in Psychology. She attended graduate school at the University of California, San Diego (UCSD), where she received her M.A. and PhD. in cognitive psychology. She was awarded a National Institute of Health post-doctoral fellowship to work on neuropsychology in the Boston Veterans Medical Center. From there she returned to San Diego for a second fellowship at

UCSD in electrophysiology. She came to NPRDC in 1987 where she works as a Personnel Research Psychologist.



GREGORY W. LEWIS was born, raised, and educated in the state of Washington. During his graduate work at Washington State University, he had extensive training in vision electrophysiology and neurophysiology. His doctoral dissertation was in the area of vision biometry using ophthalmic ultrasonography. From 1970 to 1974, Dr. Lewis fulfilled a military obligation as an Army officer in the U.S. Army Medical Research Laboratory, Fort Knox, Kentucky. Dr. Lewis has been with NPRDC

since 1974. He developed and currently heads the NPRDC Neuroscience Research Projects Office, Training Systems Department. This research is dedicated to developing techniques for improving the prediction of personnel performance by using neuroelectric and neuromagnetic waveform information. His areas of research interests include the psychophysiology of individual differences, digital processing of biological signals, and physiological correlates of brain and behavior.

BRAIN WAVE CORRELATES OF MEMORY PERFORMANCE

Diane Williams
Gregory W. Lewis

Analysis of brain waves produced by subjects engaged in a memory task provides important information about the workings of human memory. These waves provide a real-time window into mental processing. The information obtained from these waves allows us to make distinctions that are impossible to make without this information. For example, ERPs can be used to determine that subjects who may have similar performance on a certain task are actually using different cognitive strategies. Similarly, they can be used to determine that different cognitive tasks, which produce similar performance, actually rely on different mental processes. Electrophysiological techniques have been applied to cognitive paradigms and have provided important information about mental processing, which has changed our understanding of cognition. Improved assessment of cognitive strategies may ultimately provide more efficient Navy educational and training procedures.

Research Goals

Various researchers have found that the brain waves produced by the subject while studying an item can provide an index of the subject's processing and are predictive of whether that item will subsequently be recalled. This project will use brain waves to help answer questions about mental processing in two areas. First, electrophysiology will be used to help our understanding of short-term memory processing using a traditional paradigm. Second, electrophysiological data will be used to investigate the proposed existence of two different memory systems, implicit and explicit memory.

A better understanding of the workings of memory could lead to improved predictive capabilities of success in school and job performance by providing a new capability to

distinguish between people whose performance on traditional tests may be very similar.

Approach

The approach used in this research is to examine the brain waves that occur while the subjects are engaged in a standard memory task. These waves are extracted from the subject's on-going EEG by time-locking onto the information presented to the subject. When several of these waves are averaged together, the brain's response to the information emerges from the irrelevant brain activity also present in the brain waves. These time-locked waves, called event-related potentials (ERPs), have a characteristic shape that is specific to a class of stimuli presented within a certain instructional context. For example, if the subject is instructed

to listen for high tones in a series of high and low tones, when the subject hears a high tone, he or she will generate a characteristic waveform that is markedly different from the waveform generated in response to the low tones.

Short-term Memory

The memory task chosen for this research is a standard serial learning paradigm. In this task, the subject is presented with a series of eight numbers and asked to try to recall them in order. There are two effects that have emerged from this line of research: the modality effect (Corballis, 1966; Murray, 1966) and the suffix effect (Dallet, 1965). The modality effect refers to the finding that the modality of the presentation has a substantial effect on the subject's memory performance. When sets of numbers are presented visually and auditorily, performance is better for the auditorily presented lists. This effect is greatest for the last list item and extends back into the list. For the auditory list, performance is nearly perfect for the last item in the list. For the visually presented items, performance is only about 60 percent correct for the last item. This is a curious finding as the lists are informationally equivalent. Second, the suffix effect refers to the finding that if the experimenter says anything after an auditorily presented list, for example, "recall," the subject's performance in the auditory condition will be impaired. However, if the list was presented visually, the presentation of the extra word visually or auditorily does not reliably impair performance.

Despite considerable creativity on the part of researchers in this field, the modality and suffix effects are not well-understood. These effects were originally attributed to a hypothetical memory structure, called Precategorical Acoustic Store (PAS) by Crowder and Morton (1969). This memory was thought to contain auditory sensory information that had not yet been categorized for semantic content. PAS accounted for the modality effect in that it offered subjects an additional source of information about the last few items

in an auditorily presented list. When an additional word was spoken after the list, the suffix was thought to overwrite the last few items in memory, thereby decreasing performance.

While PAS provided a plausible account for these effects, a growing body of research suggests that the effects are not as circumscribed as previously thought, and consequently cannot be accounted for by such a simple theory. The suffix effect has been found in other modalities (Watkins & Watkins, 1974; Spoehr & Corin, 1978; Manning, 1980) and can be produced in the visual modality using unusual presentation techniques (Williams, 1983).

Electrophysiological data provides information that enables discriminations in performance that cannot be made using any other currently available source of data. This information will be used to determine the nature of the processing accorded stimuli in each modality and the processing given to the suffix.

Implicit and Explicit Memory

When amnesics are asked to recall a list of items, their performance is severely impaired. However, if they are provided with a list of word fragments, which are the first word that comes to mind, the words that were presented will be used to complete these word fragments at a higher than chance rate (Warrington & Weiskrantz, 1968). For example, if 'apple' was on the list, and the subject is asked to complete 'app____' to form a word, these subjects are more likely to complete that word stem to be 'apple' than they are to any other word, even though other completions would have been more likely if they hadn't seen the list that was presented to them. The subjects have no awareness of why they chose those words, but the same words that they cannot recall give evidence of having received some encoding. This pattern of data was shown to not be restricted to amnesics by Graf, Mandler, and Hayden (1982). They demonstrated that normal subjects will produce low recall and high completion if they are

asked to attend to the physical characteristics of the word rather than focusing on its meaning.

The memory assessed by using a completion task with the instructions to simply complete the word stem to form an English word was termed implicit memory (Graf & Schacter, 1985). Evidence that information is in implicit memory is inferred from the subject's behavior. In the preceding example, the amnesics, and under specific conditions, the normal subjects, showed a higher than baseline level of completion of the word stems to the words that were previously presented. This is interesting as this increase in completion performance happens in the absence of any awareness as to why these particular words were chosen as completions. Clearly, there is some memory for these words, although it is different from what we usually think of as being memory in that the subjects are unaware of these memories.

Another example of a task used to assess implicit memory is a word identification task. In this task, the subject is asked to try to read a very briefly presented word. The word is initially presented too quickly for the subject to read, and the display duration is increased until the subject can read the word. If the subjects are asked to study a list of words, then subsequently asked to read words from a display, any decrease in the display duration required to read the words from the study list compared to the display duration required to read matched words not previously studied is thought to reflect the increased activation or priming of the study list words. When this happens in the absence of being able to recall or recognize these words, this facilitation is interpreted as resulting from the storage of these words in implicit memory.

In contrast, evidence that words are in what is termed explicit memory (Graf, et al., 1985) is determined by more conventional means. The subject is simply asked to recall or recognize the words previously presented. Explicit memory is the memory typically studied in memory experiments and what is

thought of when most people think about memory.

Interestingly, the test used to assess memory does not alone determine the memory being investigated. The instructions given to the subject can change a test from being a test of implicit memory to being a test of explicit memory. For example, if a subject is instructed to complete a word fragment to be a word from the study list, the test is of explicit memory. However, if the subject is unaware of the connection between the two tasks and is just completing the word fragments with the first word that comes to mind, the test will assess implicit memory.

The distinction between these two memories is important as they are differentially affected by several variables. The frequency in the language of the words in the study time, the retention interval, and the elaborations all have different effects on implicit and explicit memory (Graf & Mandler, 1984, 1985; Graf, et al., 1982; Jacoby & Dallas, 1981).

The ERP studies in the literature of memory performance have only addressed explicit memory. Determining the relationship of the ERPs that are produced in both implicit and explicit memory paradigms might help in understanding the relationship of these two memory systems.

Progress and Plans

In FY88, the programs required to present stimuli have been written, behavioral data have been collected from Marines at Camp Pendleton, and both behavioral data and ERP data have been collected on the modality and suffix effects. This has established the feasibility of the project and provided some interesting preliminary data. For example, there has been a suggestion in the literature that the suffix acts like an additional list item, as performance for an eight-item suffixed list looks similar to performance for a nine-item list.

However, the data collected so far show that the ERPs for the last item on the list are very different from the ERPs to the suffix. Hence, although performance is similar in the two instances, the ERPs provide a way of distinguishing between situations that cannot be distinguished by solely using behavioral data.

During FY89, the research on the suffix and modality effects will be continued. In addition, research on implicit and explicit memory, which was not begun in FY88 due to late receipt of funds, will begin.

References

- Corballis, M. C. (1966). Rehearsal and decay in immediate recall of visually and aurally presented items. *Canadian Journal of Psychology*, 20, 43-51.
- Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics*, 5, 365-373.
- Dallet, K. (1965). "Primary memory": The effects of redundancy upon digit repetition. *Psychonomic Science*, 3, 237-238.
- Graf, P., & Mandler, G. (1984). Activation makes words more accessible but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior*, 23, 553-568.
- Graf, P., & Mandler, G. (1985). Direct priming across sensory modalities and work boundaries (unpublished manuscript).
- Graf, P., Mandler, G., & Hayden, P. E. (1982). Simulating amnesic symptoms in normals. *Science*, 218, 1243-1244.
- Graf, P., & Schacter, D. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 501-518.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306-340.
- Manning, S. K. (1980). Tactual and visual alphanumeric suffix effects. *Quarterly Journal of Experimental Psychology*, 32, 257-267.
- Murray, D. J. (1986). Vocalization at presentation and immediate recall, with varying recall methods. *Quarterly Journal of Experimental Psychology*, 18, 9-18.
- Spoehr, K. T., & Corin, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. *Memory and Cognition*, 6, 583-589.
- Watkins, M. J., & Watkins, O. C. (1974). A tactile suffix effect. *Memory and Cognition*, 2, 176-180.
- Williams, D. (1983). *A window on the suffix effect*. Paper presented at the annual meeting of the Psychonomic Society, San Diego, California.
- Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, 217, 972-974.

Biography



LEONARD J. TREJO was born in Mexico City, Mexico on February 24, 1955. He performed undergraduate studies at Lock Haven State College in Pennsylvania and later at the University of Oregon, receiving the B.S. (1977) degree in psychology. He performed graduate work in psychobiology at the University of Michigan and in psychology at the University of California, San Diego, with emphasis on sensation and perception. He received the M.A. (1980) and Ph.D. (1982) degrees in psychology from the University of

California for research on the neurophysiology of the pupillary light reflex and on visual sensitivity loss in hereditary retinal degeneration. From 1982 to 1984, he served as Senior Fellow in the Department of Ophthalmology at the University of Washington where he performed research on retinal toxicity, physiology of color vision, and neuroanatomy of the visual cortex. Since September 1984, he has been a Personnel Research Psychologist at NPRDC. His research interests include sensation and perception, color vision, visual pathways, cognition, and human performance.



GREGORY W. LEWIS was born, raised, and educated in the state of Washington. During his graduate work at Washington State University, he had extensive training in vision electrophysiology and neurophysiology. His doctoral dissertation was in the area of vision biometry using ophthalmic ultrasonography. From 1970 to 1974, Dr. Lewis fulfilled a military obligation as an Army officer in the U.S. Army Medical Research Laboratory, Fort Knox, Kentucky. Dr. Lewis has been with NPRDC since 1974. He developed and currently heads the

NPRDC Neuroscience Research Projects Office, Training Systems Department. This research is dedicated to developing techniques for improving the prediction of personnel performance by using neuroelectric and neuromagnetic waveform information. His areas of research interests include the psychophysiology of individual differences, digital processing of biological signals, and physiological correlates of brain and behavior.

BRAIN MECHANISMS FOR HUMAN COLOR VISION: IMPLICATIONS FOR DISPLAY SYSTEMS

Leonard J. Trejo
Gregory W. Lewis

The use of color in military displays is increasing, but the impact of color on the human operator is poorly understood. Present methods of predicting the effectiveness of color contrast in displays are based largely on behavioral threshold data, which may not be applicable to performance on dynamic visual displays. We have found that the sensitivity of individual subjects to dynamic color contrast in computer displays can be assessed by visual event-related potentials (ERPs). In most observers, we find that ERPs produced by large color differences are predicted by mathematical models based on color difference thresholds. We have also related ERPs to detection and classification performance on a task using brief signals defined by chromatic or achromatic contrast. We find that root-mean-square amplitude (RMS) measures of the P1-N1-P2 complex and the P300 component of the ERP are positively related to detection and signal classification measures of performance. These ERP-performance relationships can be understood in terms of sensory processes and selective attention, reflected by the P1-N1-P2 complex, and decision processes, reflected by the P300 component.

Problem

The interface between human operators and complex military systems is increasingly dependent on visual information displays. With the proliferation of computers as display drivers, much more information can be presented on visual displays than the operator may effectively use. Successful design of visual displays must consider sensory, perceptual, and cognitive processes of the human operator.

One focus area in visual display research is the use of color to

increase the quantity and quality of information presented to the human operator. However, the use of color in displays is proceeding without a thorough understanding of the impact of color on the human operator. In particular, most of our knowledge about human processing of color derives from behavioral research with static color displays (Burnette, 1985; Hardesty & Projector, 1973; Heglin, 1973; Meister, 1984; Merrifield & Silverstein, 1986; MIL-STD 1472C, 1981; Wagner, 1977; Wysocki & Stiles, 1982). Little is known about

the dynamics of human color processing, and even less is known about the brain mechanisms that subserve color vision.

Background

Earlier research at NPRDC has shown that measures of brain electrical responses to sensory stimuli, known as event-related potentials (ERPs), may assess unique process-related variance that relates to human performance. ERPs are very small voltage signals (microvolts) recorded from electrodes placed on the scalp that represent the response of the brain to sensory input. ERPs are usually extracted from larger ongoing electroencephalographic (EEG) activity by signal averaging. For example, the performance of individuals on a complex air defense radar simulation was correlated with the amplitude of visual ERPs produced by a series of visual probe stimuli presented during simulation performance (Trejo, Lewis, & Blankenship, in preparation). Other relationships between ERPs and performance have been demonstrated (Lewis, 1983a, 1983b).

Other research has shown that color vision is three-dimensional and that its three dimensions are subserved by three distinct brain mechanisms (Boynton, 1979). These include two chromatic (color-sensitive) mechanisms, red-green (R-G) and blue-yellow (B-Y), and one achromatic (A) or black-white mechanism. The activity of the chromatic mechanisms is thought to mediate *chromatic discrimination*, which is the ability of the visual system to discriminate colors that differ only in hue or saturation, but not in luminance. (Luminance is a measure of intensity closely related to brightness.)

The task of the display designer is complicated by the fact that chromatic

discrimination varies across stimulus conditions. Chromatic discrimination thresholds measured under one set of spatial and temporal stimulus parameters are not necessarily valid under another set of stimulus parameters. The designer must often rely on inappropriate data in specifying color contrast for information displays. Variations also exist across individuals and within an individual on a day-to-day basis and may reflect stress, fatigue, drug, or other biochemical effects. Even less is known about these individual variations than those that occur during changing stimulus conditions.

Objective

The goal of this research project is to identify physiological measures of human brain activity that carry information about chromatic discrimination and to use these measures to improve military personnel assessment and human factors engineering.

ERP measures directly related to chromatic discrimination were first reported by Riggs and Sternheim (1969). Since then, little of practical significance has been made of this important finding. One possible application of this finding is the use of ERPs for assessing the effectiveness of color contrast in information displays. Another possibility is the use of ERPs for assessing the chromatic discrimination performance of individual human subjects. Both of these issues are addressed by the basic research described in this report. We find that ERPs provide information about the effect of a color stimulus on sensory and cognitive processing by the

operator. This information, in turn, may be used in assessing the functional effects of display design features or in the selection, classification, and training of personnel who must use color-coded displays.

Approach

Our approach is to record visual ERPs produced by stimuli generated with computerized visual displays. The stimuli are presented by the method of exchange stimulation (Estévez & Spekrijse, 1982), which involves changing the color of a stimulus dynamically (over time), while holding all other parameters (e.g., size, shape, position, and texture) constant.

Progress

FY86

In FY86, a computerized system was developed to present exchange stimuli and record chromatic ERPs. ERP data were first recorded from four laboratory personnel whose color vision was tested thoroughly using clinical behavioral vision tests (Nagel anomaloscope, American Optical HRR plates, & Farnsworth-Munsell 100 Hue Test). Subsequently, chromatic ERPs were recorded from 100 military personnel during FY86 (Aug-Sep) and FY87 (Oct-Dec). These initial findings (Trejo & Lewis, 1987) demonstrated that ERPs were sensitive to pure chromatic stimulation and provided evidence of individual and day-to-day variability in chromatic ERPs.

Device-independent software was developed to allow transformation of the intensities of any RGB video monitor

into the excitation value of the human chromatic mechanisms. Supported graphics systems include the AED 512, Lexidata LEX-90, and Masscomp GA600/GA800 graphics terminals. Addition of new graphics modules requires measurement of CIE 1931 x,y chromaticity coordinates and the luminance versus input voltage functions for each of the three CRT phosphors. Several utility programs were also developed for functions such as conversion between human cone excitation values and CIE chromaticity coordinates, and calculation of CIELUV ΔE and MacAdam's Δs units of color difference from CIE chromaticity coordinates (Wyszecki & Stiles, 1982; MacAdam, 1942). All of this software was written in the "C" programming language.

FY87

In FY87, significant progress was made in the interpretation and analysis of chromatic ERPs. The number of recordings was reduced from eight to two, and the signal-to-noise ratio of the chromatic ERP was increased by approximately a factor of 10. This was accomplished by bipolar recordings of the ERP local to visual cortex and digital band-pass filtering. The results in five normal subjects demonstrated a marked similarity of properties between behavioral chromatic discrimination and the chromatic ERP. However, more information may be seen in the chromatic ERP than in behavioral measures. Specifically, ERP measures provided evidence of chromatic asymmetry in the response of the brain to the exchange of complementary colors. For example, in some subjects the exchange of green to red produced a smaller ERP than the exchange of red to green in a dynamic display. Such

direction-specific effects are difficult, if not impossible, to measure in dynamic displays using known behavioral methods. Evidence for another kind of brain asymmetry, known as lateral asymmetry, was also provided by the chromatic ERP. For example, one subject showed much larger chromatic ERPs on the right side of the head than on the left.

Results in one color deficient subject (a protanopic, or red-blind subject) demonstrated that the chromatic ERP may provide diagnostic information about color deficiency. This subject showed normal ERPs in response to exchanges containing blue-yellow contrast, but showed no significant chromatic ERPs in response to a red-green exchange.

The FY87 results were presented at the First Navy IR/IED Symposium (Trejo & Lewis, 1988).

FY88

Introduction

The aim of FY88 research was to use ERP measures to account for behavioral performance in a task requiring detection and classification of colored signals on a CRT display. This work was co-funded by a Marine Corps 6.2 research project, "Biopsychometric Assessment," aimed at developing psychophysiological predictors of on-job performance. Further analyses of the data to be described here will be performed under that project. A brief account of the major findings is presented here.

Methods

A set of 10 stimuli, which covered three critical dimensions of color contrast, were developed for presentation on a Masscomp GA600 graphics terminal. The reference point in color space (background) for these stimuli was defined by the 1931 CIE x,y chromaticity coordinate values $x=0.313, y=0.329$. This corresponds very closely to the D65 reference white of the CIE system ($x=0.313, y=0.331$). Luminance of the background was 10.3 ft-L. The two chromatic dimensions covered by the stimuli were a red-green axis and a blue-yellow (or *tritan*) axis. These axes have been used extensively in psychophysical research on color contrast sensitivity (Nagy, Eskew, & Boynton, 1987), and have been identified as cardinal directions in color space (Krauskopf, Williams, & Heeley, 1982). Along each of these axes, low and high levels of contrast were chosen in each direction (blue, yellow, red, & green) for a total of eight stimuli. Due to the coarseness of the intensity scales on the Masscomp GA600 terminal, the degree of contrast for the low and high levels along different directions were not all equal. In order to relate results obtained with color contrast to more traditional achromatic contrast (black and white), an achromatic dimension was also tested, at a high level of contrast only. This will be referred to as the achromatic condition.

The CIE 1931 x,y chromaticity coordinates, MacLeod-Boynton (1979) r,b coordinates, luminance (Y), and MacAdam's Δs color difference metrics relative to the background, and percent

luminance contrast relative to the background or modulation (M) for the 10 stimuli appear in Table 1.¹ Difficulty of a condition depends on the discriminability of the test stimuli from the background, which is inversely related to the color difference measure, Δs , in Table 1. For our stimuli, Δs corresponded better with our subjective assessment of contrast and with detection/classification performance than did the CIELUV ΔE measure. As indicated by the Δs measure, the blue-yellow contrasts should be more difficult than the red-green, and the low contrast conditions more difficult than the high contrasts. The achromatic condition is not directly comparable to the color conditions, but was chosen to be roughly equal in difficulty to the red-green high contrast condition. In order to emphasize hue differences rather than brightness differences in the color conditions, luminances of the test stimuli were chosen to be as close as

possible to that of the background. As shown in Table 1, the maximum luminance contrast among these conditions was 2.5 percent, which is near or below comparable psychophysical luminance flicker thresholds (e.g., 50 Hz thresholds, de Lange Dzn, 1958; Kelly, 1975). Luminance contrasts in the achromatic condition were about a factor of 10 higher (-15% to 25%).

¹Formulae for computation of these quantities can be found in Wyszecki and Stiles (1982). All computations were based on CIE 1931 x, y chromaticity coordinates and luminance readings from the Masscomp monitor obtained with a Thoma tristimulus colorimeter designed for use with color monitors. The error range of the chromaticity coordinates measured by this device is reported to be ± 0.005 .

Table 1
Specifications of the Color Stimuli

Color	Contrast	x	y	Y	r	b	Δs^a	% M^b
Blue	low	0.233	0.162	10.6	0.644	0.060	17.99	-1.0
Yellow	low	0.267	0.241	9.8	0.648	0.033	13.76	2.5
Red	low	0.285	0.205	10.3	0.677	0.040	41.26	0.0
Green	low	0.209	0.235	10.0	0.603	0.038	49.81	1.5
Blue	high	0.220	0.117	9.8	0.649	0.090	33.16	2.5
Yellow	high	0.270	0.308	10.4	0.634	0.022	32.21	-0.5
Red	high	0.325	0.194	10.0	0.722	0.040	87.7	1.5
Green	high	0.190	0.206	10.0	0.590	0.047	59.91	1.5
White	high	0.271	0.254	13.8	0.647	0.030	--	-15.0
Black	high	0.260	0.224	6.2	0.647	0.037	--	25.0

^aParameters were: $a = .0031$, $b = .0009$, $\theta = 1.2566$, $g_{33} = 1.0$.

^bDefined as $M = 100 (L_{\text{bkgd}} - L_{\text{stim}}) / (L_{\text{bkgd}} + L_{\text{stim}})$.

Subjects were part of a group of 106 male Marine Corps enlisted personnel stationed at Camp Pendleton. Mean age was 22.46 y ($\sigma = 2.92$). Each subject was fully briefed about the research and consented voluntarily to participate in the study. Color vision was checked using the American Optical HRR pseudo-isochromatic plates and chromatic discrimination was tested using the Farnsworth-Munsell 100-hue test. As part of the overall study, subjects first performed a short-term memory experiment (15 minutes) and three levels of a bimodal information processing task, which used achromatic checkerboard stimuli and brief, low intensity auditory tone bursts (20 minutes). Subjects then performed the color stimulus detection task in five separate trials. A single dimension of color contrast was manipulated in each trial. Trials occurred in the following order: achromatic, low contrast red-green, low contrast blue-yellow, high contrast red-green, high contrast blue-yellow.

Subjects sat and viewed the monitor at eye level from a distance of 1 m. They also held on their lap a square plastic plate on which two telegraph-style keys were mounted. As required by the preceding experiments, the left key was labeled "N" for "no" and the right key was labeled "Y" for "yes." Prior to each trial, subjects were told that a series of visual stimuli would be presented and that each stimulus in a given series could be one of two colors (e.g., red or green). They were then told to respond with a single key press to one color with the "Y" key and to the other color with the "N" key. Key presses were made with the index and middle fingers of the preferred hand (87% of the subjects were right handed). Subjects were told to respond as quickly as possible without sacrificing accuracy. The association

between colors and keys was: white-Y, black-N, red-Y, green-N, blue-Y, yellow-N. On each trial, subjects knew exactly which two colors to expect. No practice with the color stimuli was allowed. However, subjects had over 100 trials of experience in using the response keys from the prior experiments.

Each stimulus was a flash that transiently replaced a rectangular portion of the background area, which was present continuously between stimuli. Stimulus duration was a fraction (.65) of one frame of the non-interlaced video signal, or 10.8 ms. The flashed area was uniform in color and subtended 7° horizontally by 6.5° vertically (visual angle). The background was also uniform in color and subtended 14° by 9.7° . Fifty flashes were presented in each trial (25 of each color) in a pseudo-random sequence.

A nylon cap with tin electrodes (Electro-Cap International) arrayed according to the International 10-20 system (Jasper, 1958) was fitted to each subject. Active electrodes were Fz, C3, C4, Pz, O1, and O2, referenced to nose. Signals were amplified (20,000 times), band-pass filtered (0.1-100 Hz), digitized, and stored by a computer.

Silver-silver chloride electrodes were mounted above and below the right eye orbit and 2 cm lateral to the outer canthi for recording vertical and horizontal EOG. EOG gain was 10,000. These signals were subsequently used to correct the ERP recordings for volume-conducted ocular artifact (Gratton, Coles, & Donchin, 1983).

ERPs were recorded in epochs of one second duration, during which a single stimulus occurred. Recording began 125 ms before the stimulus and continued 875 ms afterward. The inter-stimulus

interval varied randomly between 1.5 and 2.5 s. If recording artifacts occurred (see below), the trials were repeated later in the sequence until 50 artifact-free trials were obtained. About 3 percent of the trials contained run-time artifacts.

For this report, analyses were restricted to the recordings from site Pz (midline parietal area). Off line, each one-second epoch was checked for artifacts (e.g., amplifier saturation, muscle artifact, movement artifact) that were not detected at run-time. This resulted in off-line rejection rates between 0 and 50 percent across subjects. In the final analysis, 40 subjects with 15 or more artifact-free epochs per stimulus were retained. Single-epoch data were then corrected for ocular artifact, digitally filtered (linear-phase, unit-gain filter, 0.5 to 25 Hz), and averaged separately for each stimulus.

Subject's responses were classified as hits, misses, false alarms, or correct rejections. For example, in the trials using red-green contrasts, pressing the "Y" key for a red flash was a hit and pressing the "N" key for a green flash was a correct rejection. Pressing the "Y" key for a green flash was a false alarm and pressing the "N" key for a red flash was a miss. The maximum time allowed for a response was 875 ms post-stimulus. Failure to respond within that period was considered to be a miss. (Response latencies were recorded; however, latency analyses will be described in a future report.) From this response classification, the hit rate, false alarm rate, and percent correct classification performance measures were computed as described by Egan (1975, pp. 8-9).

Results

Figure 1 shows two subjects' average ERP data for each stimulus from the four color trials and the achromatic trial. Both subjects were 22 years old, right handed (self report) and right-eye dominant (sighting test), reported having good acuity (no need for glasses), and had normal color vision as shown by perfect performance on the HRR pseudo-isochromatic plates. Both subjects also had normal color discrimination as shown by the 100-hue test; however, subject #71 had an error score of 24, which is nearly in the range of superior ability (cutoff is 16), whereas subject #73 had an error score of 90, which is nearly in the range of poor ability (cutoff is 100). Subject #71 rated himself as being 100 percent alert, reported 8.5 hours of sleep the night before, and had not consumed coffee, tea, or other stimulants. Subject #73 rated himself as being 75 percent alert, reported 6.5 hours of sleep the night before, and had consumed a caffeine-containing beverage and had smoked tobacco before reporting for the experiment. Both subjects were tested between 12:00 and 3:00 pm.

The average ERPs in Figure 1 are arranged vertically according to the difficulty of the color condition with the most difficult, BYLO (low contrast blue-yellow), first and the easiest, RGHI (high contrast red-green), last. The ACHR (achromatic) condition appears at the bottom. Within each condition there is a separate curve for each of the two colors that were presented. The lightweight line marks the ERPs for the "target" colors--those associated with the "Y" key (blue, red, white). The heavy line marks the ERPs for the "non-target" colors--those associated with the "N" key (yellow, green, black). Time zero on the abscissa marks the stimulus

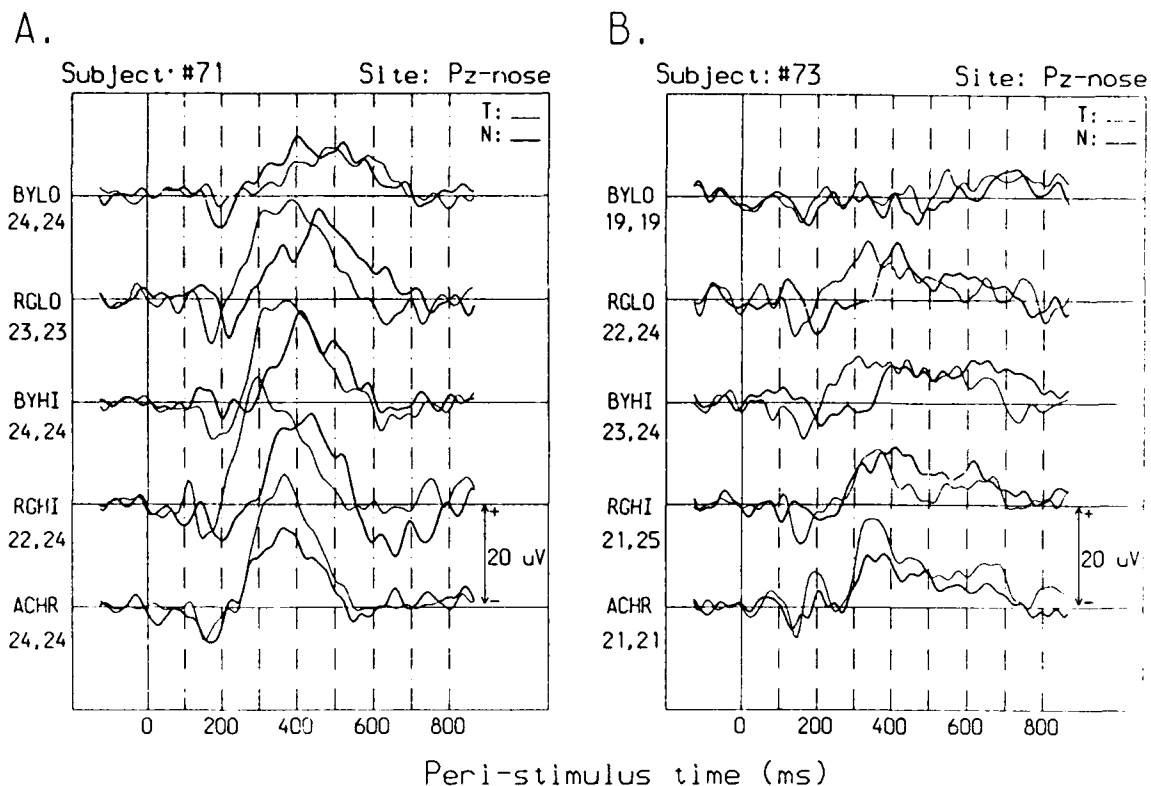


Figure 1. Average ERPs obtained in the four color conditions and the achromatic condition in two subjects. Numbers of single epochs included in each average appear on the ordinate. Stimulus occurred at time zero on the abscissa. Recording site was Pz, the midline parietal site, and reference electrode was on the nose. Positive voltages are plotted upwards. Light traces are for target ("Yes") colors: blue, red, and white. Heavy traces are for non-target ("No") colors: yellow, green, and black.

onset time. Positive voltages are plotted upward. The number of artifact-free single epochs used to compute each average ERP appears on the ordinate for target and non-target colors respectively.

The features of the average ERP can be divided into two salient segments, which can be seen in the examples of Figure 1. The first segment contains three peaks. First there is a positive peak with a maximum between 100 and 150 ms post-stimulus. This is most evident in

the ERPs for the RGHI condition. Next, there is a negative peak that reaches a minimum between about 150 to 225 ms post-stimulus. This is evident in every ERP. Another positive peak sometimes follows the negative peak, reaching a maximum between 200 and 250 ms post-stimulus. This is most evident in the achromatic ERPs. The second segment consists of a large positive peak that reaches a maximum between 275 and 475 ms post-stimulus and may appear to contain multiple peaks within that period. This segment appeared to vary

substantially across subjects and conditions in relation to stimulus discriminability and task performance.

In order to quantify these two segments of the ERP, we computed the root-mean-square voltage in two adjacent time windows. The first window extended from 50 to 250 ms post-stimulus and will be referred to as the N1-RMS. The second window extended from 250 to 500 ms post-stimulus and will be referred to as the P3-RMS. Similar measures have

previously been shown to index selective attention and cognitive workload (Blankenship, Trejo, & Lewis, 1988).

We evaluated the effect of our stimulus conditions on these ERP measures using two separate repeated-measures analyses of variance. Independent factors were stimulus type (target or non-target) and color condition (ACHR, RGLO, RGHI, BYLO, BYHI). The results are shown in Table 2. Both the N1-RMS and the P3-RMS were significantly dependent on stimulus type and color condition.

Table 2

Analysis of Variance for N1-RMS and P3-RMS Measures

Source	SS	df	F ¹
A. N1-RMS Measure (50-250 ms post-stimulus)			
Condition (C)	60.69	4	9.74
CxS	242.91	156	
Color (CL)	26.65	1	21.09
CLxS	46.29	39	
CxCL	20.61	4	5.54
CxCLxS	145.13	156	
B. P3-RMS Measure (250-500 ms post-stimulus)			
Condition (C)	246.49	4	12.72
CxS	755.61	156	
Color (CL)	137.57	1	40.89
CLxS	131.20	39	
CxCL	57.91	4	4.63
CxCLxS	487.83	156	

¹ $p_a < 0.001$ for all these effects.

In general, the RMS voltage for both segments was positively related to the ease of the required perceptual discrimination. This is illustrated graphically in Figure 2. The N1-RMS was nearly constant for the BYLO, RGLO, and BYHI conditions (mean $2.88 \mu\text{V}$) but was about 27 percent higher for the RGHI and ACHR conditions. On average, the P3-RMS

was twice as high as the N1-RMS and provided finer discrimination among conditions, being higher for red-green contrasts than for blue-yellow contrasts and higher also for high contrasts than for low contrasts.

For both the N1-RMS and the P3-RMS, the interaction between stimulus color (target versus non-target) and condition was significant. This is

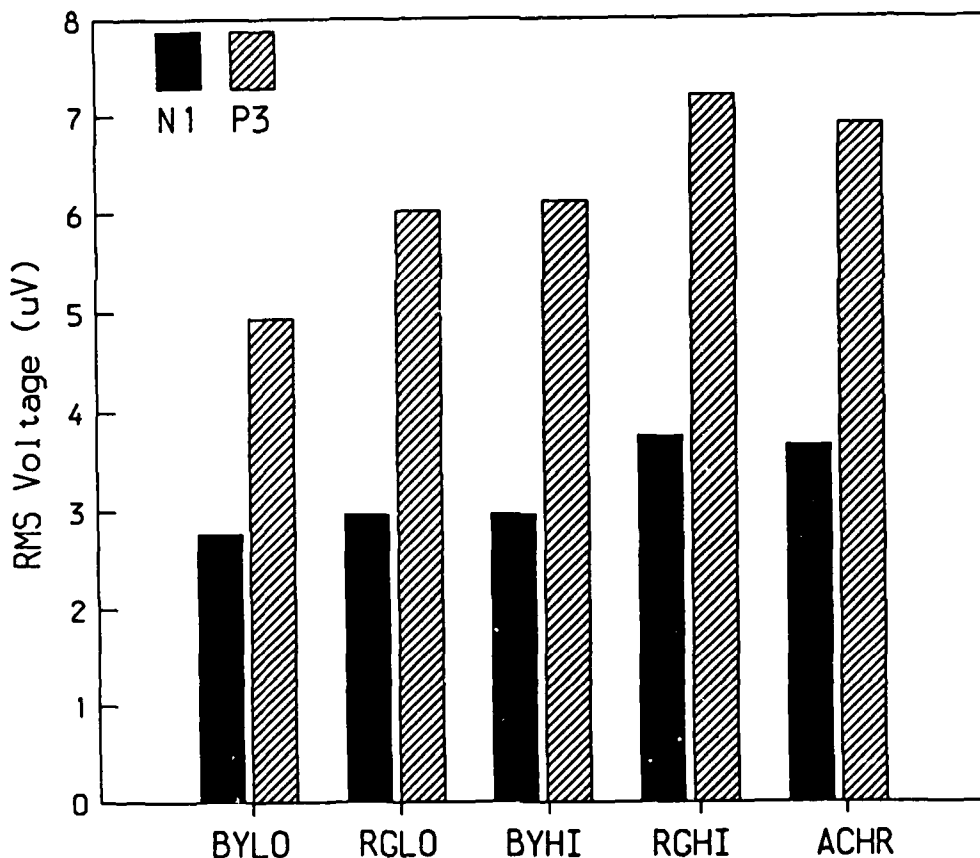


Figure 2. Mean values of N1-RMS and P3-RMS voltages in each condition obtained in our 40-subject sample. N1-RMS represents amplitude in the ERP between 50-250 ms. P3-RMS represents ERP amplitude between 250-500 ms. Color condition is on the abscissa, RMS voltage on the ordinate.

illustrated in Figure 3. Target colors (blue, red, white) produced higher RMS values than non-target colors (yellow, green, black) in all conditions except the low contrast blue-yellow condition, where target and non-target RMS values were about equal. In part, these differences may be due to greater contrast between target colors and the background than between non-target colors and the background. In every condition except RGLO, the color with the largest value of Δs produced the largest RMS values. Furthermore, in the ACHR condition, the white target had greater absolute luminance contrast from the background than did the black non-target, which also agreed with the higher RMS values for white. However, in the RGLO condition, the red target color had a lower Δs value than the non-target green, yet still produced higher RMS values than the green. Alternatively, the associations of the word "Yes" with the targets and "No" with the non-targets may have resulted in selective

attention or response bias effects that enhanced the processing of the target stimuli. This could also have the effect of elevating N1 and P3 component amplitudes. Unfortunately, there was not enough time to counterbalance the associations of responses with colors to decide between these alternatives.

We performed a multiple correlation and regression analysis to examine the relationship between the detection and classification performance measures and the ERP measures. For the target stimuli, the appropriate detection measure is the hit rate, HR, (probability of pressing the "Y" key when a target stimulus occurred) whereas for the non-target stimuli, the appropriate detection measure is the correct rejection rate, CR, (probability of pressing the "N" key given that a non-target stimulus occurred). For both stimuli in a condition, the appropriate classification measure is the probability of a correct decision, PC. Typically PC is computed from estimates of the

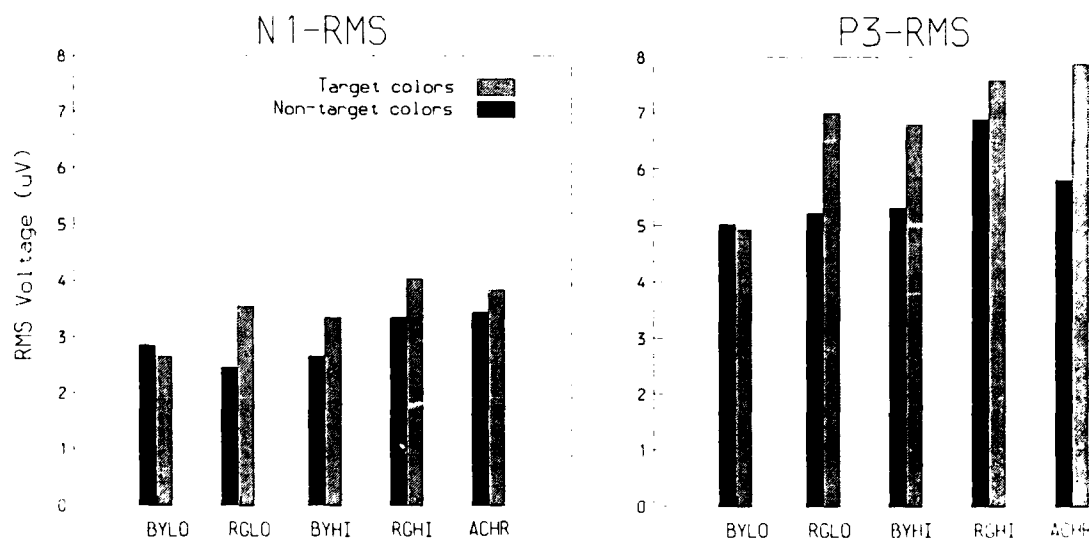


Figure 3. Mean values of N1-RMS and P3-RMS voltages for target and non-target colors as a function of color condition obtained in our 40-subject sample. Graphing conventions are the same as in Figure 2.

probabilities of hits and correct rejections (Egan, 1975). However, in our study some subjects made responses that were forbidden by the test paradigm. These included pressing both keys in response to a stimulus and failing to press a key within the response interval of 875 ms post-stimulus. These responses were classified as "nonspecific errors." We computed the error rate, ERR, as the ratio of total non-specific errors to the total number of stimuli presented. An overall performance measure, PC-E, was computed as the difference between the probability of a correct response, given that an allowable response was made, and twice the non-specific error rate ($PC-E = PC - 2ERR$).

Correlations were computed between HR, CR, PC-E, ERR and the target and non-target RMS measures separately for each stimulus condition. Surprisingly, no significant correlations between behavioral performance on the task and RMS measures were observed for the achromatic condition. In each color condition, the correlation matrix indicated a direct relationship between RMS amplitude measures and HR, CR, and PC-E. An inverse relationship held between ERR and the RMS measures. These relationships were clearer for the difficult stimulus conditions than for the easier conditions. In addition, the correlation coefficients were much larger in magnitude for the P3-RMS than for the N1-RMS.

Using the trends indicated by the pattern of correlations seen within conditions, we devised a single multiple regression equation to describe the relationship between detection/classification performance and the ERP measures considered as predictors. Due to its dependence on HR, CR, and ERR, PC-E was selected as the best overall measure of detection/classification

performance. Based on the magnitude of the coefficients within conditions, the P3-RMS for targets (P3T) and the P3-RMS for non-targets (P3N) were chosen as predictors. The regression model was $PC - E = .071P3T + .068P3N - .669$.

The multiple r value of this model was 0.521, which accounts for 27 percent of the variance in performance ($F_{2,157} = 29.163, p < .0001$). For comparison, a model including the target and non-target N1-RMS values was also tested, but the addition of the N1-RMS measures increased the variance accounted for by only 0.74 percent.

Discussion

The data clearly showed that ERPs produced by chromatic and achromatic stimuli in a detection and classification task contain information about sensory and cognitive processing of those stimuli. First, it was shown that the amplitude of the ERP increased monotonically with the degree of color contrast present in the stimuli. Thus, the ERP may serve as a gauge of the effectiveness of color contrast in conveying useful information to the operator of a color display. Both the N1-RMS and the P3-RMS exhibited this property, but the relationship was better defined for the P3-RMS. Second, it was shown that colors associated with "Yes" responses produced larger N1-RMS and P3-RMS values than colors associated with "No" responses. This effect did not appear to be entirely accounted for by differences in contrast presented by the target and non-target stimuli. It is possible that this effect is related to the cognitive factors, which predisposed subjects to attend selectively to the target color. Greater attention to targets should, in general, produce larger ERP amplitudes, particularly for components in the range covered by our

N1-RMS measure (Harter & Aine, 1984). Although we did not measure components per se, nor did we compare ERPs for colors when they were targets and non-targets, the larger N1-RMS amplitudes for all target stimuli we observed is consistent with increased attention to targets.

The data also showed the P3-RMS measures for target and non-target stimuli were directly related to detection and classification performance in the color stimulus conditions. As much as 27 percent of the variance in overall detection/classification performance was accounted for by this measure alone. However, the strength of this relationship appeared to be inversely related to the difficulty of the color discriminations required for good performance. For example, the weakest correlations between performance measures and the P3-RMS were observed in the high contrast red-green condition, and no significant correlations were observed in the (easy) achromatic condition. This effect may reflect the presence of a limiting ceiling in the P300-RMS. It may be that a gradation in P300 amplitude exists only in the range of stimulus discriminations that produce measurable numbers of performance errors. For discriminations that are essentially error-free, P300 amplitude may be constant. This idea is supported by the observation that performance was excellent in both the achromatic and the high contrast red-green conditions, with probabilities of correct detection (PC) of .92 and .87 respectively. In contrast, PC values in the BYLO, RGLO, and BYHI conditions were .26.

Conclusion

This research project has provided hardware, software, experimental

methods and data that have directly augmented the ongoing research efforts of the Neuroscience Projects Office at NPRDC. In particular, the capability to design and run experiments involving color stimuli on electronic displays for performance assessment has been greatly enhanced. Finally, valuable data have been collected, which document the relationship of the ERP to human color processing and human performance with color displays.

Recommendations

We recommend that research on color vision and ERPs be continued at NPRDC within the context of 6.2 (exploratory development) efforts aimed at specific application areas. For example, the design of future color-coded displays may benefit from physiological data, which describe the perceptual processing of color. In addition, the color ERP may prove useful in selecting or training personnel who must interact with color coded displays.

Future basic research in color ERPs should also be performed. Factors that merit future study include different regions of color space, spatial and temporal stimulus properties, refinement of ERP measures, evoked magnetic field measures, and analyses of individual differences.

References

- Blankenship, M. H., Trejo, L. J., & Lewis, G. W. (1988). *Brain activity and tactical decision-making: IV. Event related potentials as indices of selective attention and cognitive workload* (NPRDC Technical Note 89-6). San Diego: Navy Personnel Research and Development Center.
- Boynton, R. M. (1979). *Human color vision*. New York: Holt, Rinehard, & Winston.

- Burnette, C. A. (1985). *Uses of color in complex information displays* (HFOSL Technical Note 71-85-04). San Diego: Navy Personnel Research and Development Center.
- de Lange Dzn, H. (1958). Research into the dynamic nature of the human fovea cortex systems with intermittent and modulated light: I. Attenuation characteristics with white and colored light. *Journal of the Optical Society of America*, 48, 777-784.
- Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.
- Estévez, O., & Spekreijse, H. (1982). The "silent substitution" method in visual research. *Vision Research*, 22, 681-691.
- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55, 468-484.
- Hardesty, G. K. C., & Projector, T. H. (1973). *NAVSHIPS display illumination design guide. Section I: Introduction to light and color*. Annapolis, MD: Naval Ship Research and Development Center.
- Harter, M. R., & Aine, C. J. (1984). Brain mechanisms of visual selective attention. In R. Parasuraman and D. R. Davies (Eds.), *Varieties of attention*. Orlando, FL: Academic Press.
- Heglin, H. J. (1973). *NAVSHIPS display illumination design guide. Section II: Human factors* (Technical Document 223). San Diego: Naval Electronics Laboratory.
- Jasper, H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10, 211-224.
- Kelly, D. H. (1975). Luminous and chromatic flickering patterns have opposite effects. *Science*, 188, 371-372.
- Lewis, G. W. (1983a). Event-related electrical and magnetic activity: Toward predicting on-job performance. *International Journal of Neuroscience*, 18, 159-182.
- Lewis, G. (1983b). *Bioelectric predictors of personnel performance: A review of relevant research at the Navy Personnel Research and Development Center* (NPRDC Technical Report 84-3). San Diego: Navy Personnel Research and Development Center.
- Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research*, 22, 1123-1131.
- MacAdam, D. L. (1942). Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America*, 33, 18.
- MacLeod, D. I. A., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, 69, 1183-1186.
- Meister, D. (1984). Human engineering data base for design and selection of cathode ray tube and other display systems (NPRDC Technical Report 84-51). San Diego: Navy Personnel Research and Development Center.

- Merrifield, R. M., & Silverstein, L. D. (1986). *The development and evaluation of color systems for airborne applications: Fundamental visual, perceptual, and display systems considerations* (NADC Rep. No. 86011-60). Warminster, PA: Naval Air Development Center.
- MIL-STD-1472C (1981). *Human engineering design criteria for military systems, equipment, and facilities*. Washington, D C: U. S. Department of Defense.
- Nagy, A. L., Eskew, R. T., & Boynton, R. M. (1987). Analysis of color discrimination ellipses in a cone excitation space. *Journal of the Optical Society of America (A)*, 4, 756-768.
- Riggs, L. A., & Sternheim, C. E. (1969). Human retinal and occipital potentials evoked by changes in the wavelength of the stimulating light. *Journal of the Optical Society of America*, 59, 635-640.
- Trejo, L. J., & Lewis, G. W. (1987). Brain mechanisms for human color vision. In A. M. Crawford & M. H. Metcalfe (Eds.), *FY86 Independent Research/Independent Exploratory Development annual report* (NPRDC Special Report 87-1). San Diego: Navy Personnel Research and Development Center.
- Trejo, L. J., Lewis, G. W., & Blankenship, M. H. (In preparation). *Brain activity during tactical decision-making: III. Relationships between physiology, simulation performance, and job performance*.
- Trejo, L. J. & Lewis, G. W. (1988). Sensitivity to hue differences measured by visual evoked potentials. In D. Eggleston (Ed.), *First Navy Independent Research/Independent Exploratory Development Symposium* (CPIA Publication 492), 309-316. Laurel, MD: Chemical Propulsion Information Agency.
- Wagner, D. W. (1977). *Color coding: An annotated bibliography* (NWC Technical Publication 5922). China Lake, CA: Naval Weapons Center.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae*. New York: John Wiley & Sons.

Biography



ROBERT F. MORRISON is the Head of the Career Development Systems Division at NPRDC where he has worked since 1976. He was born in Minnesota and raised in Iowa acquiring a B.S. (1952) in general science and an M.S. (1956) in applied psychology from Iowa State University. His PhD. (1961) in industrial psychology was received from Purdue University. Dr. Morrison worked in human resource management, emphasizing career development, for Mobil Corporation, the Mead Corporation, and Martin -Baltimore. He has

headed the personnel research activity for Sun Company and his own consulting firm as well as teaching in the School of Management at the University of Toronto. His areas of research interest are management identification and the career development of managers, professionals, and scientists. His professional publications include a book, book chapters, and professional and scientific articles. Dr. Morrison's major research won the James McKeen Cattell Award for Research Design.

EXPERIENCED-BASED CAREER DEVELOPMENT

Robert F. Morrison

Human resource specialists need to be able to design systematically patterns of assignment that lead to the development of effective performance in positions many years after the career development process is begun. The objective of this research is to identify the steps and the time it takes an individual to master a single assignment and use this as a component in a life-span model of experiential learning. This effort provides an initial description (model) of the factors that influence how long it takes individuals to develop expertise within a specific assignment. When the research is completed, the Navy will have an algorithm to add learning time and performance level factors to the present methodology used to establish tour length. At this time manpower and permanent-change-of-station (PCS) cost factors are the major factors considered.

Background and Problems

One hidden assumption with the growth of huge, formal education and training programs is that all efficient learning must occur in a structured (classroom-like) systematic way. Preliminary research on managerial positions challenges that assumption and indicates that the majority of learning occurs as a result of work experience (Brousseau, 1984; Campbell (personal communication), 4 January 1985; Hall & Fukami, 1979; Kanarick (personal communication), 3 April 1981; Lombardo, 1982; Morgan, Hall, & Martier, 1979; Vineberg & Taylor, 1972). While a model and propositions covering an entire career has been proposed (Morrison & Hock, 1986), the detailed attributes of its components were not

adequately defined. This definition is imperative to the adequate explication of the career development process.

Since the Navy moved from pursuing its primary warfare mission in the mid-seventies to a peacetime status, the demands for its personnel, especially officers, to perform effectively in a wide variety of tasks and situations have increased markedly. A program to encourage unrestricted line (URL) officers in the development of a secondary skill (subspecialty) has floundered yet culminated in the introduction of a material professional community in 1986. In 1981, the Surface Warfare Commanders Conference focused on junior officers to increase their technical skills. In 1985, Congress imposed the requirements that all must serve in a joint billet in order

to be eligible for promotion to flag (0-7).

This plethora of demands has forced policymakers to shorten billet and command tours until they are frequently less than 18 months. Such policies have been designed using manpower flow models without considering their effect on the officers' performance and career development. The fleet's personnel readiness and the effectiveness with which support activities perform are affected directly by the opportunity that officers have to develop the capability to learn the requisite knowledge and skill of each billet and to develop them to a level of mastery. Tour lengths that are too short do not provide the opportunity to develop while ones that are too long make inefficient use of the officer force and may lower the officers' motivation to perform at a high level or learn new tasks/jobs.

Objective

The broad objective of this research is to develop a generic model describing the factors that influence how long it takes an individual to develop an expert-level of skill in performing work. The specific objective is to develop, qualitatively test, and modify a preliminary model of the learning that occurs while the incumbent is in a leadership position.

General Approach

A literature search was used to identify the steps that a leader goes through to learn the job to a point of mastery and the parameters (individual, job, and organizational) that contribute to the individual's entry state, what is learned, how it is learned, and how quickly it is learned. The information derived from the literature review was used to form an initial model of the experiential learning process and the factors influencing it. Repeated interviews with 26 surface warfare department heads and 8 executive

officers from 9 surface combatant and amphibious ships were used to collect data that would provide a qualitative test of the initial model. As a result of the interviews, the initial model was revised and research was designed to test a situationally specific model of experiential learning on the population of surface warfare department heads. At this stage, the model hypothesizes four major sets of factors that influence learning of the job via experience. The sets are individual differences, job, internal environment, and the external environment as shown in Figure 1 and described in detail in *An experience-based learning model: A pilot study* (Morrison & Brantner, in review).

Plans

Upon completion of the situationally specific population test of the experiential learning model on surface warfare department heads, the work will transition into exploratory development (6.2). Within the exploratory development phase, generalizability of the model can be evaluated and, if necessary, further modification can be done. Later, specific decision rules/algorithms can be defined to aid Navy manpower policymakers in making decisions concerning tour length using developmental and performance factors in addition to manpower flow and PCS cost variables.

Expected Benefit

The Navy spends millions of dollars annually on PCS moves that are based primarily on manning and PCS cost considerations. By adding individual career development and performance factors to the decision process, the readiness of the fleet and the effectiveness with which PCS dollars are used will improve. If training proves to be a significant factor in the model, the model will provide a basis for evaluating training programs.

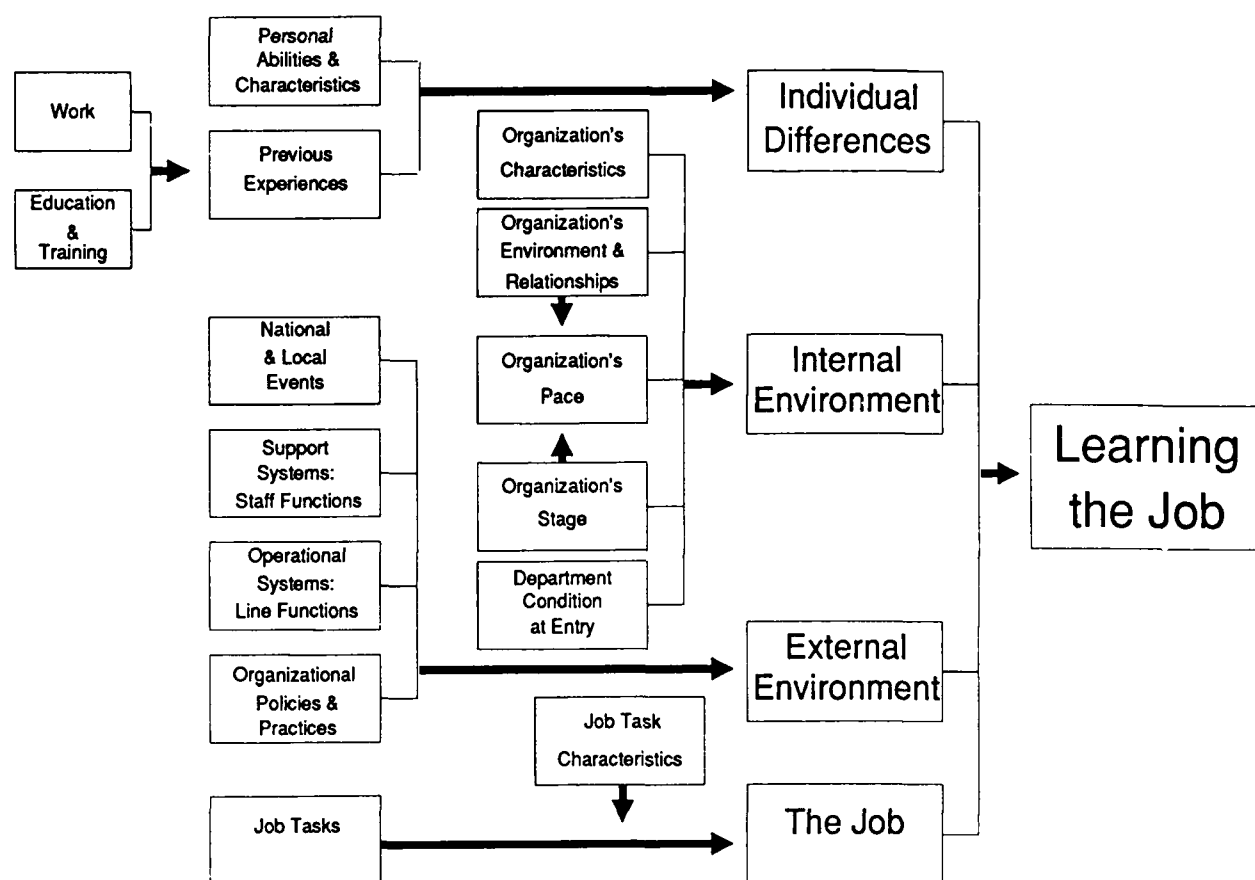


Figure 1. Job learning model.

References

- Brousseau, K. R. (1984). Job-person dynamics and career development. *Research in Personnel and Human Resources Management*, 2, 125-154.
- Hall, D. T., & Fukami, C. V. (1979). Organization design and adult learning. *Research in Organizational Behavior*, 2, 125-167.
- Lombardo, M. M. (1982). How do leaders get to lead? *Issues and Observations*, 2(1), 1-4.
- Morgan, M., Hall, D. T., & Martier, A. (1979). Career development strategies in industry: Where are we and where should we be? *Personnel*, 56(2), 13-31.
- Morrison, R. F., & Hock, R. R. (1986). Career building: Learning from cumulative work experience. In D. T. Hall (Ed.), *Career development in organizations*, (236-273). San Francisco: Jossey-Bass.
- Vineberg, R., & Taylor, E. N. (1972). *Performance in four Army jobs by men at different aptitude (AFQT) levels: 3. The relationship of AFQT and job experience to job performance* (HumRRO TR 72-22). Alexandria, VA: Human Resources Research Organization.

Biography



DONALD BAMBER was a Research Psychologist in the Human Factors Department at NPRDC until that department was transferred to the Naval Ocean Systems Center (NOSC) in mid FY88. He now holds the position of Scientist in the Information and Decision Management Branch at NOSC. He has a B.S. in Mathematics from M.I.T. and a PhD. in Experimental Psychology from Stanford. Before coming to NPRDC, he was a Research Psychologist at the Veterans Administration Medical Center in St. Cloud, MN

where he developed new mathematical methods for testing models of memory impairment. His current research interests include human decision making, computerized decision aiding, elicitation of knowledge from experts, and mathematical modeling. He is a member of the *Psychonomic Society*, the *Human Factors Society*, the *Society for Judgment and Decision Making*, the *Society for Mathematical Psychology*, the *American Mathematical Society*, the *American Statistical Association*, the *Cognitive Science Society*, and the *American Association for Artificial Intelligence*. He serves on the editorial board of the *Journal of Mathematical Psychology*.

HOW TO ELICIT KNOWLEDGE FROM EXPERTS

Donald Bamber

The initial step in building an expert system is to elicit the relevant knowledge from an expert. Unfortunately, experts often have difficulty recalling things they know. In particular, experts often recall generalizations but cannot recall all the exceptions. In order to more reliably elicit from experts their knowledge of generalizations and exceptions, I began by reviewing the literature in artificial intelligence on nonmonotonic reasoning. While this literature provided me some insight into the difficulties of reasoning about generalizations and exceptions, I concluded that artificial intelligence researchers had sufficiently different goals from mine and that I would need to take a different approach from theirs. I generated a corpus of generalizations and exceptions, sorted them into categories, and extracted what appeared to be the principles underlying them. Using these principles as a foundation, I am currently working to develop a "logic" of generalizations and exceptions. I expect that this "logic" will have implications for how knowledge of generalizations and exceptions should be elicited from experts when building expert systems.

Background

Expert systems are computer programs that give advice to nonexperts so that they can perform tasks that normally require expertise. Typically, these programs are developed after interviewing experts and observing them at work. This process is called *knowledge elicitation*. The programs are then designed to follow the same principles and procedures that the experts use.

Expert systems have potential for widespread use in the Navy. However, that potential will not be realized unless expert systems can be built in a reliable and cost-effective manner. Currently, one of the most difficult steps in building an expert

system is the elicitation of the necessary knowledge from an expert. This step is time-consuming and, worse yet, prone to error. Expert systems containing incomplete or inaccurate knowledge are potentially dangerous; in critical situations they may fail catastrophically. Thus, it is most important that any knowledge elicitation procedure used to develop an expert system should yield knowledge that is as complete and accurate as possible.

Part of the lore of expert system development is that it is relatively easy for experts to state generalizations but relatively difficult for them to state the exceptions to those generalizations. For example, an expert may state that, under certain circumstances,

usually the best course of action is to do X. The expert also says that there are exceptions to this rule. When asked to enumerate the exceptions, the expert lists a few but then states that there are additional exceptions that he or she can't think of at the moment.

This fact poses a problem for expert system developers. If experts occasionally forget exceptions to generalizations that they tell to an expert system developer, then the resulting expert system will sometimes make errors because it doesn't know all the exceptions that it should.

Objective

The goal of this project is to develop a theory of how generalizations and exceptions are represented in human memory, how they are retrieved, and how people process them once they have been retrieved. It is desirable to have such a theory because it would provide guidance as to how best to elicit knowledge about generalizations and exceptions from experts.

Approach

I began by reviewing the literature in the field of artificial intelligence (AI) that deals with nonmonotonic reasoning.

To understand what is meant by the term *nonmonotonic reasoning*, it is necessary to understand what is meant by its opposite, namely, *monotonic reasoning*. Consider standard logic. For any set of axioms, there are certain propositions that can be proven from those axioms. Adding new axioms to the axiom set makes it possible to prove additional propositions that could not have been proven using only the original axioms. Moreover, adding new axioms

can never make a formerly provable proposition unprovable. So, because increasing the set of axioms can only increase the set of provable propositions, standard logic is said to be monotonic.

Human reasoning frequently does not have this property of monotonicity. Consider the following example frequently found in the AI literature. If a person is told that Tweety is a bird, the person is liable to use his or her knowledge that birds fly to infer that Tweety can fly. Now, had the person been told that, not only is Tweety a bird, Tweety is an ostrich, then the person would have used his or her knowledge that ostriches are an exception to the generalization that birds fly and would have concluded that Tweety can't fly. So, in this case, giving the person additional information (Tweety is an ostrich) would cause the person not to make an inference (Tweety can fly) that he or she would have made otherwise. A reasoner, whether human or computer, is said to be nonmonotonic if it will sometimes not make an inference that it would have made if given less information. Clearly, much human reasoning about generalizations and exceptions is nonmonotonic.

Many AI researchers have sought to develop theories of nonmonotonic reasoning with the goal of employing nonmonotonic reasoning in AI programs. I reviewed much of the AI literature on nonmonotonic reasoning hoping that it would give some insight into human reasoning. Now, unless one has a strong background in mathematical logic, the nonmonotonic reasoning literature can be a challenge to read. It took me a while to penetrate beyond the mathematical formalism to understand the basic ideas that guide this research. What I gained from reading this literature is an

appreciation for the difficulties that must be overcome by a nonmonotonic reasoner. My view now is that, while the AI researchers' proposed methods of nonmonotonic reasoning may be good for AI programs, they are not the methods employed by humans.

Consider, for example, the technique of circumscription (McCarthy, 1980, 1986). A reasoner that employs circumscription will not "jump to conclusions" the way that humans sometimes do. Suppose a circumscriptive reasoner is told the generalization *birds can fly* together with the exception *ostriches can't fly*. Essentially what the circumscriptive reasoner does is to form a conjecture to the effect that the *only* exceptions to the generalization *birds can fly* are the exceptions that it already knows. Thus, the circumscriptive reasoner forms the conjecture *all birds can fly except ostriches*. Once it has formed this conjecture, the circumscriptive reasoner does its reasoning in accordance with standard logic. Thus, if the circumscriptive reasoner is told that Tweety is a bird but is not told whether or not Tweety is an ostrich, the reasoner will not jump to the conclusion that Tweety can fly. Not knowing whether Tweety is an ostrich prevents the circumscriptive reasoner from reaching any conclusion regarding whether or not Tweety can fly.

Next, consider Reiter's (1980) default logic. In this logic, generalizations may be expressed as defaults. A typical default is: *Any bird can fly unless it can be proven that it can't fly*. Exceptions are expressed as ordinary logical statements. A typical exception is: *Ostriches can't fly*. Suppose that a reasoner using default logic is given the previous generalization and exception. Suppose the reasoner is also told that

Tweety is a bird but is given no other information about Tweety. Being unable to prove that Tweety can't fly, the reasoner applies the default *any bird can fly unless it can be proven that it can't fly* and concludes that Tweety can fly. On the other hand, suppose that, instead of merely being told that Tweety is a bird, the reasoner were told that Tweety is an ostrich. Then the reasoner would apply the exception *ostriches can't fly* to conclude that Tweety can't fly. Moreover, having a proof in hand that Tweety can't fly, the default *any bird can fly unless it can be proven that it can't fly* wouldn't be applicable.

It seems unlikely to me that humans employ a default logic such as Reiter's. To see this, consider the Tweety example once more. In order to apply the default *any bird can fly unless it can be proven that it can't fly*, in essence what must be done is to demonstrate that there exists no proof that Tweety can't fly. In the example just given, there was only one fact known about Tweety, namely, that Tweety is a bird. Thus, it was easy to see that there was no way to prove that Tweety can't fly. However, in a more complex example with many more known facts, it could be very difficult to demonstrate that there existed no proof that Tweety can't fly. Yet, unless this could be done, it would not be permissible to apply the default *any bird can fly unless it can be proven that it can't fly*. Thus, within default logic, proving one statement may require demonstrating that there is no proof of some other statement. To demonstrate that a proof does not exist can be computationally intensive. For this reason, I doubt that humans use default logic.

In general, much of the AI literature on nonmonotonic reasoning systems has been concerned with the conclusions

that an "ideal reasoner" would reach. An ideal reasoner is one that makes neither errors of commission nor errors of omission in its reasoning. In other words, an ideal reasoner never infers conclusions that are not justified and never fails to infer a conclusion that is justified. A strong motivation for studying ideal reasoning is that it provides a normative standard against which to judge the performance of a computer program. However, by definition, an ideal reasoner has superhuman reasoning capabilities. Thus, while ideal reasoning might provide a normative standard for human reasoning, it does not constitute a model of human reasoning.

In my opinion, a more fruitful approach to modeling human reasoning is to construct models in which the reasoner may make errors. These errors would occur because the reasoner would sometimes fail to consider some of the relevant information. However, the reasoner could recover from such errors when he or she eventually did take into account the previously unconsidered information. For example, suppose the reasoner were informed of the facts *Tweety is a bird* and *Tweety is an ostrich*. The reasoner might use the fact *Tweety is a bird* and the generalization *birds can fly* to erroneously conclude *Tweety can fly*. However, as he or she continued reasoning, the reasoner would use the fact *Tweety is an ostrich* together with the exception *ostriches can't fly* to conclude *Tweety can't fly* and to overturn the original conclusion *Tweety can fly*. This style of reasoning seems similar in spirit at least to Doyle's (1979) truth maintenance system.

Reasoning in which errors are made and then recovered from seems a particularly plausible style of reasoning for humans given that, in life, not all

relevant information is received simultaneously. For example, suppose that Tweety is an ostrich but that a person is informed only that Tweety is a bird. This person is liable to infer the erroneous conclusion *Tweety can fly*. If later the person is informed that Tweety is an ostrich, then the person will need to recover from the previous error and conclude *Tweety can't fly*. In fact, in everyday life, people do seem to recover from errors induced by the sequential arrival of information. It seems plausible that people may make reasoning errors and recover from them not only because items of information may be received sequentially but also because items of information may be received simultaneously but considered sequentially.

At this point, I felt that the AI literature had given me some understanding of the difficulties of nonmonotonic reasoning. However, I also felt that the AI systems of nonmonotonic reasoning were not descriptive of human reasoning about generalizations and exceptions. I decided that, because the goals of the AI researchers were different from mine, I would need to take a different approach to the problem.

My next step was to generate a corpus of generalizations and exceptions. I did this by introspection. I generated statements that I believed were typically true and then tried to think of exceptions to those statements. Next, I looked through the generalizations and exceptions and sorted them into groups that appeared to be governed by common principles.

Generalizations and exceptions divide into two main groups: those dealing with matters of fact and those dealing with matters of choice. Both

these types of generalizations and exceptions are important in expert systems. The former are employed when making inferences about the world, while the latter are employed when deciding which of several alternative actions is preferable.

The exceptions to generalizations regarding matters of fact divide into two subgroups. The first subgroup is implicit exceptions. Belief in these exceptions derive from other knowledge. For example, suppose Fred is a bird who has had surgery severing the nerves to his wings. Because I believe that birds cannot fly unless they flap their wings and that Fred's surgery prevents him from flapping his wings, I conclude that Fred is an exception to the generalization *birds can fly*. In other words, my belief in this exception is implicit in my beliefs about bird flight, wing flapping, and the function of nerves. A second subgroup of exceptions are explicit exceptions. Belief in these exceptions is not derived from other knowledge. For example, my belief that ostriches are an exception to the generalization *birds can fly* is not derived from any other beliefs. This is an explicit exception.

Generalizations regarding matters of choice take the form: To achieve goal *G*, the best action to take is *A*. A typical example is that usually the best way for me to get to work is to drive my car. This generalization is reasonably sound in that it is correct well over 90 percent of the time. However, there are a number of different exceptions.

Exceptions to generalizations regarding matters of choice divide into several subgroups: (1) I might not take an action to achieve a goal if that would prevent me from achieving another goal. For example, I will not drive my car to

work if it is necessary to have my car serviced. (2) I will not take an action if that action will not achieve its goal. For example, I will not drive my car to work if the car is in the garage and the garage door is broken and cannot be opened. (3) I might not take an action to achieve a goal if an alternative action is available. For example, I will not drive my car to work if someone offers me a ride. (4) I might not take an action if that action would have negative consequences. For example, if I discover that the wrong type of oil has been put in my car and I believe that it would severely damage the engine to drive the car, then I will not drive to work.

Having studied a wide range of generalizations and exceptions, I am now convinced that these are not arbitrary. Rather, human thinking about generalizations and exceptions is governed by principles.

Ongoing Work

I am now working to develop a set of rules for reasoning about generalizations and exceptions. Such a "logic" should satisfy certain criteria. First, its description should be mathematically rigorous. Second, it should allow reasoning in which errors are made but then recovered from. Third, it should explain a wide variety of generalizations, exceptions, exceptions to exceptions, etc. pertaining both to matters of fact and to matters of choice. Fourth, it should make testable predictions about human reasoning.

Expected Benefit

We will be better able to elicit from experts their knowledge of generalizations and exceptions when we have a theory of how people reason about generalizations and exceptions.

This will enable us to develop expert systems more reliably and more efficiently.

References

Doyle, J. (1979). A truth maintenance system. *Artificial Intelligence*, 12, 231-272.

McCarthy, J. (1980). Circumscription-- A form of nonmonotonic reasoning. *Artificial Intelligence*, 13, 27-39.

McCarthy, J. (1986). Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28, 89-116.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13, 81-132.

Biography



MERYL SUE BAKER was born in Brooklyn, New York on November 21, 1950. She earned a B.A. degree with a dual major in Psychology and Education from the University of Toledo, a masters degree in Educational Psychology from Michigan State University, and a PhD. in Instructional Psychology from Florida State University. Dr. Baker was employed by Courseware, Inc. for 2 years as an Instructional Psychologist prior to coming to NPRDC in late 1978. At NPRDC, Dr. Baker was project director for the R&D of the Navy's Job

Oriented Basic Skills program before becoming head of the Instructional Science Division in 1984. In 1986, Dr. Baker moved to her current position as head of NPRDC's Training Systems Applications Division.

READING COMPREHENSION STRATEGIES

Meryl Sue Baker

The purpose of this effort was to (1) test the effectiveness of reading comprehension lessons with Navy recruits in an operational setting and (2) examine the net payoff to the Navy of improving reading comprehension skills of recruits.

The Armed Forces Qualifying Test (AFQT) and Nelson Denny Reading Test scores of 52 students awaiting instruction at the Naval Submarine School, Groton, CT were collected. The mean of the AFQT and Nelson Denny scores served as the basis for forming matched experimental and control groups. Reading strategies guides were distributed with less than an hour of verbal explanation/instruction. Written instructions are contained within the self-instructional guides. This procedure was adopted to meet the Navy need of limiting the number of classroom contact hours devoted to basic skills instruction. Subject submarine school performance data were collected and analyzed. Results showed no differences between experimental and control groups in terms of school performance. Results were interpreted to indicate that recruits will likely not review strategies on their own time, and unless significant time is invested in teaching reading strategies, they will likely not prove useful.

Background and Problem

The reading deficiencies of Navy recruits have been well documented. Though many Navy dollars have been invested in reading skills programs, little payoff has been realized. The problem with most Navy remedial reading programs, as well as most civilian remedial programs designed for adolescents, is that they focus on phonics and vocabulary. These areas often are not the primary deficient capabilities of adult readers. An assessment of the decoding and vocabulary ability of adult poor readers often indicates such skills to be at a functional level.

A review of the few experimental studies available in the area of remedial reading for adolescents/adults clearly indicates that an approach to remedial reading that employs total language use with emphasis on comprehension, the area in which most adults experience the greatest deficiencies, affords the greatest gains in reading.

In response to this requirement, 15 comprehension strategies were identified. These strategies individually were found to either positively affect previously poor readers, or were found to be employed by good readers. These 15 strategies

were developed into self-instructional lessonware to be used by Navy personnel experiencing reading comprehension difficulties. The self-paced reading comprehension module contains three lessons: pre-reading, comprehension, and retention. Accompanying the module is a guide for its use.

Objective

The purpose of the present research was to test the effectiveness of the aforementioned reading comprehension lessons with Navy recruits in an operational setting.

General Approach

Implementation took place at the Navy Submarine School, Groton, CT. To be of maximum use to the Navy, remedial instruction should not increase the length of time the student is in class. This can be achieved by requiring students to learn material on their off-duty hours or by demonstrating that a remedial program decreases follow-on school contact hours sufficiently as to not increase overall time at a school. With the Submarine School at Groton, we attempted the former. Subjects were students who were waived into the Submarine School and were awaiting the start of classes. Due to their "waivered" status their Armed Forces Qualifying Test (AFQT) scores represented a wide range (25 to 99).

Subjects were divided into groups with comparable AFQT scores. Each subject in the experimental group was given the guide for the self-instructional materials as well as the materials themselves. They were acquainted with the materials during a 1 hour introductory session where it was emphasized that they were not to share the materials with anyone. Control group subjects received no materials.

Data were collected on Submarine School performance of each of the experimental and control groups.

Results

Analysis of Submarine School performance (first segment) showed no differences between the experimental and control groups. Analyses were also conducted on the differences between low and high AFQT groups and again no differences were found. Findings are being interpreted to indicate that it is unlikely that subjects actually read the materials during their off-duty hours. It could also be that the materials were ineffective or produced at a reading grade level (RGL) too high for the audience. Pending further research, which actually controls usage of the materials, it is difficult to comment on their effectiveness. However, they were designed employing an instructional model that has proved effective in the past in a variety of different subject areas. As for the RGL of the materials, sampling of the materials indicates the RGL to be well within the range of the high AFQT group.

Plans

Ideas and strategies explored during this effort will transition into the 603720 project, "Prerequisite Skills Enhancement Program." This is a broad based program whose purpose is to reduce or eliminate the Navy "A" school basic/prerequisite skills deficiencies of Navy technical school students. It is anticipated that reading difficulties will be identified as one of the most critical and severe deficiencies.

Expected Benefit

The expected benefit of reading research is to improve reading abilities such that technical school attrition is decreased, overall training costs are lowered, and Navy manpower requirements are satisfied.

Biography



IOSIF KRASS is an Operations Research Analyst in the Manpower Systems Department at NPRDC. His research specialties are large scale optimization and dynamic control systems. For the last 3 years he has been developing and applying advanced techniques in operations research, computer science, and control theory to solve complex problems in Navy enlisted personnel assignment and rotation. Educated and trained in the Soviet Union as an operations research analyst, he started his research work at the Institute of

Mathematics, Siberian Branch of Academy of Science, Novosibirsk, USSR (1967-1978). After immigrating to the U.S., he was an Associate Professor at Southern Illinois and Kansas Universities (1979-1981). He also worked as a systems and programming consultant for Control Data Corporation in Connecticut (1981-1985). He is a member of the *Operations Research Society of America*. He has over 20 years of research experience and he has authored or co-authored 2 books, 28 journal articles, and 19 professional papers.

OPTIMAL CONTROL THEORY FOR A SYSTEM OF QUASI-LINEAR DIFFERENCE EQUATIONS

Iosif Krass

The Navy's sea/shore rotation policy is based on fixed tour lengths. This leads to underachievement of personnel distribution goals and reflects false commitments to individual enlisted members. The Navy wants to test policies which result in better alignment of a dynamic personnel inventory with a set of dynamic manpower requirements. This effort develops a method that allows flexible tour lengths and represents sea/shore rotation as a dynamic control system.

Background and Problem

The Navy's sea/shore rotation policy must support readiness, personnel stability, and career attractiveness over a protracted period of time (5 to 8 years). It is recognized that significant changes in both inventory and requirements will occur during the duration of the policy. The current system is based on fixed tour lengths, but has evolved into a system of exceptions to the fixed tours. As a result, the Navy cannot determine the impact of a fixed tour length policy on the future distribution of personnel inventories or the cost of rotation.

The Navy is considering a major change in policy in rotating its enlisted personnel between sea duty and shore duty. Instead of the traditional fixed lengths of assignments (tour lengths) at sea and shore, the Navy wants to test more flexible policies. Enlisted rotation managers must respond to both long-range policy goals and near-term fluctuations in personnel vacancies. A methodology is needed which

allows tour lengths to be flexible in order to meet the Navy's personnel readiness needs.

Objective

The objective of this research is to develop an improved method by which the Navy can make sea/shore rotation decisions to align a dynamic personnel inventory with the dynamic billet authorizations.

General Approach

Much progress has been made over the past several years toward solving large and complicated dynamic economic systems by means of Von Neumann modeling techniques. Using this approach, the sea/shore rotation problem was formulated as a dynamic control system of quasi-linear difference equations with controls. This technique was then used to determine the best trajectory for the system's hierarchical objectives. In technical terms, the calculation of this trajectory can be set as an optimal solution of a network with

side constraints. A computer solution was obtained using a network with side constraints code developed at Southern Methodist University together with a subsequent rounding routine specially designed for this kind of system.

Results

A dynamic modeling framework was developed and tested for the Navy's sea/shore rotation problem. A paper documenting the formulation and test results entitled, "An application of dynamic modeling to the sea shore rotation planning problem in the Navy," was accepted for publication in *Computer and Mathematics with*

Applications, an international journal. A computer implementation of the model was also transitioned to a 6.4 effort, sea shore rotation management system.

Expected Benefit

The benefits of this research will be seen when the new Navy Enlisted Personnel Rotation System is implemented. The new system is expected to result in increased readiness through more flexible rotation policy, reduced turbulence, and improved retention, improved morale due to consistent application of policy, and well justified and defended PCS budgets.

Biography



JAMES P. BOYLE is a statistician in the Manpower Systems Department at NPRDC. He was trained in Mathematics and Statistics at the University of Missouri, receiving a PhD. in Statistics in 1984. His research interests include empirical Bayes estimation and force management forecasting models. He is presently a member of the *American Statistical Association*.

LOSS FORECASTING WITH EMPIRICAL BAYES ESTIMATORS

James P. Boyle

Manpower planners within the Navy and Marine Corps require accurate personnel loss forecasts to generate executable personnel plans. Traditionally, the method of least squares has been a popular method of estimating parameters in loss forecasting models. In recent years, a group of statistical techniques known as empirical Bayes methods have emerged as competitors to the least squares approach. This research effort was designed to test the hypothesis that empirical Bayes estimators generate more accurate loss forecasts than standard least squares estimators.

Background and Problem

Military manpower systems are "vacancy driven." New personnel are recruited and current personnel are promoted when vacancies are created by various types of personnel losses. Plans to achieve targeted end-strengths and satisfy budget constraints are good only if projected losses come close to actual losses. Effective force planning depends heavily on accurate loss forecasts.

While there is a wide variety of loss forecasting models in use, the models can be grouped broadly into two categories--reflecting quite different viewpoints. The first group consists of time series models that assume future losses are entirely a function of past loss behavior. The second group, comprised of econometric models, explain future loss behavior in terms of exogenous variables (e.g., military compensation, civilian employment conditions).

Although different in spirit, the two approaches share a common characteristic. Both contain parameters that must be estimated before any forecasts can be obtained. A popular method of estimating these parameters has been least squares.

However, in recent years, the popularity of least squares has been challenged by a set of techniques known as "empirical Bayes" methods. The emergence of these "empirical Bayes" methods is due largely to a series of papers published by Bradley Efron and Carl Morris from 1972 to 1977. These papers contain both theoretical and empirical results recommending empirical Bayes estimators over least squares estimators. A book by G. G. Judge and M. E. Bock (1978), which develops empirical Bayes estimators in the General Linear Model, and a recent article by George Casella (1985) also confirm the superiority of empirical Bayes estimators.

Objective

The objective of this research is to test the hypothesis that empirical Bayes estimators of parameters in loss forecasting models lead to better forecasts than those based on least squares estimators.

Approach

A time series regression model was applied to quarterly Marine Corps loss data. The data exhibited quarterly variation in both level and trend, leading to the consideration of multiple intercept and slope parameters. Several empirical Bayes estimators of these parameters were developed. Forecasts based on the empirical Bayes estimates were compared to forecasts based on the usual squares estimates.

Results

The test data consisted of 24 quarterly observations (FY81 through FY86) on end-of-active-service loss rates for the entire enlisted Marine Corps and 3 occupational fields (OccF): OccF3 (Infantry), OccF25 (Operational Communications), and OccF30 (Supply Administration and Operations). For each of the four series, two models were considered. A mean square forecast error criterion was adopted to rank the various estimation techniques. In seven of the eight cases, all empirical Bayes estimates generated better forecasts than the least squares estimates of the model parameters. Details of the FY88 effort can be found in an NPRDC technical Note 88-54 entitled *An Empirical Bayes Approach to Forecasting Marine Corps Enlisted Personnel Loss Rates*.

Also in FY88, an enlarged class of empirical Bayes estimators was derived.

This work was presented at the 1988 annual meeting of the *American Statistical Association (ASA)*. A paper describing this research is to appear in the ASA 1988 proceedings of the *Business and Economics Statistics* section.

Plans

In FY89, an expanded class of time series loss forecasting models will be applied to enlisted Marine Corps loss data. Again, forecasts based on empirical Bayes estimators will be compared to those based on standard estimators. Additionally, work will commence in assessing the performance of empirical Bayes techniques as applied to econometric models of loss behavior. Results from this project will transition into an existing 6.2 project (Marine Corps force management forecasting), and two existing 6.3 projects (Marine Corps enlisted planning system and the Navy's distributable inventory management information system).

Expected Benefit

Force management decisions impact billions of dollars in manpower appropriations. Since personnel loss forecasting is critical to successful force management, this effort commands considerable leverage. Benefits are expected to take the form of personnel plans that achieve skill and experience levels required to sustain readiness and avoid budget crises.

References

- Carter, G., & Rolph, J. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *Journal of the American Statistical Association*, 69, 880-885.

- Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, 39, 83-87.
- Efron B., & Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators. Part II: The empirical Bayes case. *Journal of the American Statistical Association*, 67, 130-139.
- Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors: An empirical Bayes approach. *Journal of the American Statistical Association*, 68, 117-130.
- Efron, B., & Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70, 311-319.
- Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, 236(5), 119-127.
- Fay, R. E. III, & Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.
- Judge, G. G., & Bock, M. E. (1978). *The statistical implications of pre-test and Stein-Rule estimators in econometrics*, Amsterdam: North-Holland.
- See page B-3 for FY88 publication.

Biography



J. BRADFORD (BRAD) SYMPSON is a Personnel Research Psychologist in the Testing Systems Department. Brad is one of a handful of individuals in this country who have developed a high level of expertise in Item Response Theory (IRT). IRT is a new approach to mental testing that is revolutionizing the way tests are designed, administered, and scored. After completing advanced graduate work in psychometrics (the theory of mental testing) at the University of Minnesota, Brad spent 2 years at Educational

Testing Service working with Dr. Frederic M. Lord, one of the initial developers of IRT. Brad came to NPRDC in 1981 to work on the Joint Services Computerized Adaptive Testing (CAT) Project. Brad has developed several new statistical procedures and IRT models for use in personnel testing. He is a member of the *American Educational Research Association*, the *American Statistical Association*, the *Military Testing Association*, the *National Council on Measurement in Education*, the *Personnel Testing Council of San Diego*, and the *Psychometric Society*. In addition to serving as a manuscript reviewer for several journals published by these professional organizations, Brad is currently the President of the *Personnel Testing Council of San Diego*.

MODELS FOR CALIBRATING MULTIPLE-CHOICE ITEMS

J. Bradford Simpson

Dichotomous (right/wrong) scoring of multiple-choice test questions does not distinguish among the various wrong answers chosen by examinees. Wrong answers can supply valuable information about an examinee's capabilities. In this project, new item-response models and a polychotomous item-scoring procedure were developed. Application of this new technology to military selection, classification, and achievement testing can improve personnel decisions.

Problem

In current applications of multiple-choice questions to mental testing (e.g., in the *Armed Services Vocational Aptitude Battery* and in training courses), examinee responses are scored as either correct or incorrect. This dichotomous item-scoring procedure does not distinguish among the various incorrect answers that examinees select. Information about an examinee's level of knowledge that could be extracted from wrong answers is lost.

Objective

The objective of this project was to develop new psychometric (psychological measurement) procedures that would extract additional information about an examinee's level of knowledge from the examinee's wrong answers to test questions. It was anticipated that such procedures would increase the reliability of test scores, thus supporting improved personnel decisions in military selection, classification, and training.

Progress

This project was previously funded under the NPRDC Independent Research (IR) Program. In FY88, the project was funded under the Independent Exploratory Development (IED) Program. Funding in FY88 was \$46K, all expended in-house. Following are the major accomplishments of the project:

1. Several polychotomous item-response models were developed and tried out using available test data (Simpson 1983, 1986a, 1986b, 1987b).
2. A new procedure for computing scoring weights for all the response options of a multiple-choice item was developed (Simpson, 1984, 1987a, 1988a).
3. A new family of statistical distribution functions and a computer program that fits this distribution function to sets of test scores was developed (Simpson & France, 1984).

4. Research results indicate that the new technology developed in this project increases test reliability by an amount that is equivalent to a 25 percent increase in test length (Simpson, 1986b; Simpson & Haladyna, 1988).

5. The Principal Investigator planned and coordinated a 2-hour symposium on polychotomous item-scoring procedures that was presented at the 1988 meeting of the American Educational Research Association. Participants included the Principal Investigator and faculty-members from the Universities of Arizona, Chicago, Colorado, Illinois, Kansas, and Maryland. Invited Chairperson for the symposium was Dr. Charles E. Davis, a Scientific Officer at the Office of Naval Research.

Benefits

1. Empirical results (Simpson, 1986b; Simpson & Haladyna, 1988) show that the polychotomous item scoring methods developed in this research do provide additional information about examinee ability. Application of these methods will allow us to shorten mental tests by about 20 percent, without sacrificing test reliability.

2. The best polychotomous model developed in this research has "fit" every test item to which it was applied. Thus, if this model is implemented, more of the test questions that are written can be used.

3. Our procedures allow test developers to identify test questions and response alternatives that are especially good or especially poor indicators of ability or knowledge, and aid in determining the nature of the processes that underlie examinee responses.

All of these benefits will serve to improve personnel decisions that are made in military selection, classification, and training.

Presentations Prior to FY88

Simpson, J. B. (1983, June). A new item response theory model for calibrating multiple-choice items. Paper presented at the meeting of the *Psychometric Society*, Los Angeles, CA.

Simpson, J. B. (1984, October). Principal component analysis of polychotomous item responses. Talk given at the *Office of Naval Research Conference on Model-based Psychological Measurement*, Princeton, NJ.

Simpson, J. B. (1986a, April). Some item response functions obtained in polychotomous item analysis. Talk given at the *Office of Naval Research Conference on Model-based Psychological Measurement*, Gatlinburg, TN.

Simpson, J. B. (1986b, August). Extracting information from wrong answers in computerized adaptive testing. In B. F. Green (Chair), *New developments in computerized adaptive testing*. Symposium conducted at the meeting of the *American Psychological Association*, Washington, DC.

Simpson, J. B. (1987a, May). Extracting information from wrong answers in personnel testing. Invited talk presented to members of the *Personnel Testing Council of San Diego*, San Diego, CA.

Sympson, J. B. (1987b, June).

Application of a polychotomous IRT model to adaptive mental testing.

Paper presented at the *Office of Naval Research Conference on Model-based Psychological Measurement*, Columbia, SC.

Sympson, J. B., & France, S. L. (1984, October). A family of distribution functions. Paper presented at the *Office of Naval Research Conference on Model-based Psychological Measurement*, Princeton, NJ.

See page B-1 for 1988 presentations.

Biography



DELBERT M. NEBEKER is a Personnel Research Psychologist at NPRDC and a former Associate Professor of Management at San Diego State University. In addition to his duties at NPRDC, he finds time to serve as an Adjunct Professor at the California School of Professional Psychology and is President of the consulting firm, Performance Matters. He is the author of numerous professional journal publications and government technical reports. He has served on the

Editorial Board for *Organizational Behavior and Human Decision Processes* for over 10 years and has received many awards and honors. These include the 1979 Military Psychology Award from the *American Psychological Association* and an Exemplary Practices in Federal Productivity award from the *United States Office of Personnel Management* in 1981. He is listed in *Who's Who in the West*. He received his PhD. in Psychology from the University of Washington and a B.S. in Sociology and Psychology at Brigham Young University.



PAUL H. De YOUNG is a Personnel Research Psychologist in the Organization Systems Department. Mr. De Young received a B.A. in Psychology and a B.S. in Biology from the University of California, Irvine. Prior to entering his graduate training, Mr. De Young worked as a Behavioral Specialist in the private sector. He received a masters degree in Industrial/Organizational Psychology in 1986 at the California School of Professional Psychology and is currently a doctoral candidate. As a graduate student,

Mr. De Young served as a research assistant in the Organizational Systems Laboratory (OSSLAB) where he collected his dissertation data. In addition, he worked as an Intern at a local Medical Center at which he designed and implemented several survey-guided feedback programs. He currently is involved in a project to evaluate the implementation of Total Quality Management in four Navy organizations. Mr. De Young is a member of the *Academy of Management* and the *American Psychological Association*.

Biography



B. CHARLES TATUM was born in Champaign, Illinois on October 29, 1947. He earned advanced degrees (M.A., PhD.) at the University of New Mexico in General Experimental Psychology. Before coming to NPRDC, Dr. Tatum was an associate professor and chairman of the psychology department at Cornell College in Iowa. While in Iowa, he was also employed as a marketing and advertising consultant and did research and post doctoral studies at the University of Iowa and the

University of Michigan. Since coming to NPRDC, Dr. Tatum has worked on several projects in the areas of organizational design and command effectiveness. He is currently involved in a project to implement "gain sharing" (an employee work-incentive program) in four Navy organizations. Dr. Tatum is a member of the *American Psychological Association* and the *Academy of Management*. He also serves as an Adjunct Professor at San Diego State University and the California School of Professional Psychology. He has published 11 papers (journal articles and technical reports) and given 16 presentations at professional meetings.

GROUP SIZE AND MEMBER APPROVAL OF REWARD PLANS IN A GAIN SHARING SYSTEM: EFFECTS ON INDIVIDUAL PERFORMANCE

Delbert M. Nebeker
Paul H. DeYoung
B. Charles Tatum

The current budget deficit is one of the most critical problems facing the U.S. Federal Government. We are faced with the challenge of developing new management methods and enhancing old ones to improve productivity. Gain Sharing systems have shown the potential to make substantial improvements in productivity, yet relatively little is known about how they can be most effective. This article discusses a study conducted in a simulated organization which examines some of the critical dimensions of a Gain Sharing system. It was found that the effects of Gain Sharing on employee productivity are moderated by work group size, the degree of co-worker approval of the system, and worker ability. Workers with relatively higher levels of ability and who perceived their co-workers to highly endorse the Gain Sharing system achieved greater performance improvements than those with low ability and who were in low endorsement conditions. In addition, a significant increase in worker performance under the Gain Sharing system was found when employees worked in relatively small groups of six people with high co-worker endorsement of the system. The results from this research provides valuable information in helping the Navy develop techniques for increasing industrial productivity and efficiency.

Problem

The budget deficit is one of the most serious and difficult problems currently facing the U.S. Federal Government. The 1987 Budget Deficit Reduction Act (often referred to as the Gramm-Rudman-Hollings Act) was enacted by Congress to deal with the problem. President Reagan, realizing that just cutting the budget was not enough, signed Executive Order 12637 on April 27, 1988. The Order establishes a government-wide requirement to improve the quality, timeliness, and efficiency of services provided by the Federal Government.

The scope of this Executive Order is broad and encompasses all of the

military, including the U.S. Navy. The Order calls for a 3 percent annual average increase in productivity by all Executive Departments and Agencies. This increase is more than double the historical average rate of increase in the Department of Defense. The Navy will be faced with serious manpower and budgetary constraints in the near future as the full effect of these actions is felt. Clearly, if the goals of the Executive Order are to be met, and the fiscal crisis resolved, major changes must be made in the way Navy organizations are managed. Management is thus faced with the challenge of finding new methods, and enhancing old ones, to improve productivity.

Background

One method for improving productivity is to increase employee motivation through performance-based incentive systems. There are a variety of such systems that range from individual systems (such as wage incentive systems common in manufacturing settings) to organization-wide reward systems (such as Improshare, the Scanlon Plan, and the Rucker Plan).

Group-based systems are often referred to as Gain Sharing plans. Specifically, Gain Sharing plans offer monetary bonuses (a share of the gains from increased productivity) to all employees for productivity improvement. Usually, the increased productivity is the result of highly involved group efforts in finding better ways to operate. The bonuses are often distributed in equal shares to all employees.

There has been a heightened interest recently in these Gain Sharing systems, both in government and private industry (Mohr, et al., 1985; Ross & Hauck, 1984; Bullock & Bullock, 1982; Hammerstone, 1987). Organizations have turned to Gain Sharing because these systems have resulted in impressive productivity gains. A recent survey of 1,598 organizations conducted by the American Productivity Center reported that firms with Gain Sharing systems in operation more than 5 years averaged almost 29 percent savings in work-force costs (cited by Boyett, 1987).

Most of what is known about Gain Sharing systems comes from case studies that describe a single organization's efforts to install a plan. Additional research focuses on some of the situational factors that favor the

organization's success in a sample of ongoing plans. Relatively little, however, is known about why Gain Sharing systems work (Schuster, 1984; Bullock & Bullock, 1982). This is because the studies of research and reward systems and Gain Sharing systems are very limited, and of the studies done, most do not meet rigorous methodological standards (Lawler, 1985). Consequently, there is a great need for controlled studies that focus on how group rewards motivate individuals and how the effectiveness of these group incentive programs can be maximized.

Nebeker, Neuberger, and Hulton (1983) identified several critical system design dimensions that may influence the effectiveness of reward systems. In addition, Davis (1969) has identified environmental and individual difference variables that may influence individual performance in groups. Some of the variables noted by Davis include the size of the group, group cohesiveness, and group norms.

Typically, Gain Sharing systems are developed without much attention paid to most of these dimensions. Yet, there is much evidence to suggest that varying the levels of these dimensions can have a significant effect on an individual's job performance. For example, Marriot (1949) found that the size of the work group can have an effect on work group output. He found that as the size of the group increased, there was a concomitant decrease in the output of the work group. Similarly, Latane, Williams, and Harkins (1979) found that on a simple manual dexterity task, individual performance decreased with increasing group size (a phenomenon they call "social loafing").

In contrast to the above research findings, there are data of a more

practical, applied nature that show Gain Sharing systems are successful irrespective of the size of the organization (White, 1979). Thus, there is a discrepancy between the research evidence that group size is negatively related to individual performance, and the practical evidence that Gain Sharing is successful regardless of the number of individuals involved. One possible explanation for this discrepancy may be that under a Gain Sharing system, individuals are less likely to "loaf" because they believe that other members of the work unit "approve" of the Gain Sharing system (i.e., they support and endorse the plan, and are eager to make it work). The present research was designed to resolve the discrepancy in the evidence by addressing two related questions: (1) With increasing work group size, does individual performance decline under conditions where group members disapprove of the Gain Sharing system? (2) Under conditions of high group member approval, does individual performance remain constant regardless of the size of the work group?

Approach

We conducted the research in the Organizational Systems Simulation Laboratory (OSSLAB) at the Navy Personnel Research and Development Center to establish a relatively high degree of experimental control while maintaining a high degree of fidelity to actual work settings. Seventy-two subjects were recruited through a local temporary employment agency and were hired as "Data Base Operators" to enter and maintain a computerized data base. The employees worked 5 days a week, 4 hours a day, over a period of 2 weeks in a simulated organizational setting.

The research design was a $2 \times 3 \times 2$ mixed factorial design. The first independent variable was a within-subject manipulation that included a no Gain Sharing condition (first-week baseline period) and a Gain Sharing condition (second-week). The Gain Sharing treatment involved paying bonuses to the employees when group performance exceeded a standard. The standard was based on the average level of group performance during the baseline period. The next two independent variables were between-subject factors; (1) work group size, and (2) approval of the Gain Sharing system. The study used three levels of work group size (6, 12, and 24 persons). The actual size of the group was always 4 persons, but the subjects were led to believe that they were working in larger groups (the remaining 2, 8, or 20 members were phantom workers whom the real workers believed were working in a different location). The study included two levels of Gain Sharing system approval (high and low). Individuals were given a questionnaire and asked to rate the degree to which they supported the Gain Sharing plan. On the following day, the workers were given contrived feedback on how their group responded to the questionnaire. Half of the groups were given feedback on the questionnaire results that indicated the majority of their co-workers highly approved of the plan; the other half were given feedback that the majority of their co-workers disapproved of the plan.

The subjects were paid \$5.15 per hour during the baseline period and then, as noted above, received a group reward during the remainder of the study. We treated the employees in much the same fashion as workers are treated in organizations with real Gain Sharing systems (e.g., employees were

encouraged to improve their performance, the monetary bonus was based on a 50% sharing of the costs for productivity gains). We controlled for possible interactions between the group's actual performance and the amount of the group bonus by holding the level of group performance constant across all conditions. We accomplished this control by creating an artificial standard and reporting fictitious levels of group performance. This false feedback allowed us to predetermine the amount of Gain Sharing bonus pay given to employees. The true levels of individual performance remained free to vary and were accurately reported to the subjects, but the group data were fictitious.

Data were continuously and automatically collected by micro-computer work stations that measured such things as time spent in individual tasks, the number of keystrokes per hour, and the time and frequency of rest breaks. The dependent variable of major concern was productivity, measured in keystrokes per hour.

Results

Daily average performance scores were chosen as the dependent variable in a moderated multiple regression. The days chosen were day 3 and 4 in the baseline or non-Gain Sharing condition and days 8 and 9 in the Gain Sharing condition. These days were chosen to minimize any learning effects by using only those days where performance had stabilized within each condition. Days 5 and 10 were excluded as a precaution against any unusual effects that might have occurred on the day that questionnaires were administered. The variables entered in the regression analysis were: (1) Performance on day 2 as an individual difference variable

(ability) and covariate; (2) Group size; (3) Gain Sharing; (4) Group approval of Gain Sharing system feedback; and (5) all higher order interactions.

The results show a large and significant portion of the performance variance accounted for by the independent variables. Over 85 percent of the performance variance was accounted for by the experimental variables and their interactions ($R = .928$; $\text{Adj } R^2 = .851$; $F = 76.413$; $df = 21, 257$; $p < .00001$). Only two main effects are significantly related to performance. They are individual ability level and the Gain Sharing treatment. Neither group size nor the group's approval of the Gain Sharing system are, by themselves, significantly related to performance. However, a significant amount of additional performance variance is accounted for by two higher order interactions. The first is an interaction between ability, Gain Sharing, and approval of Gain Sharing. The nature of this interaction is such that the Gain Sharing and approval treatments are most effective on higher ability workers. The second significant interaction is between Gain Sharing, size, and approval. The effect of this interaction is to increase the performance of those operators in the six person groups when under Gain Sharing and with high group approval for the Gain Sharing system. These effects are most easily seen when plotted graphically. Figures 1, 2, and 3 show the performance of high and low ability workers with and without the Gain Sharing system when members of approving and disapproving groups. Figure 1 shows the results for six person groups, Figure 2 for 12 person groups, and Figure 3 for 24 person groups. As can be seen from the data plots, Gain Sharing systems increase performance for almost all workers.

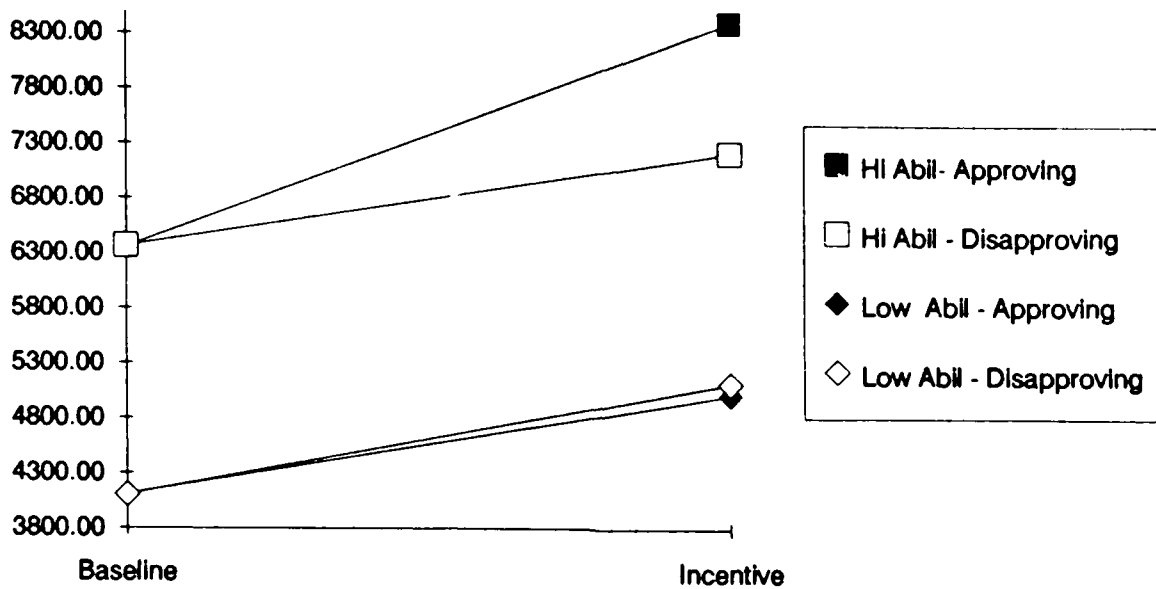


Figure 1. Performance of high and low ability operators in six person groups which approve and disapprove of reward system.

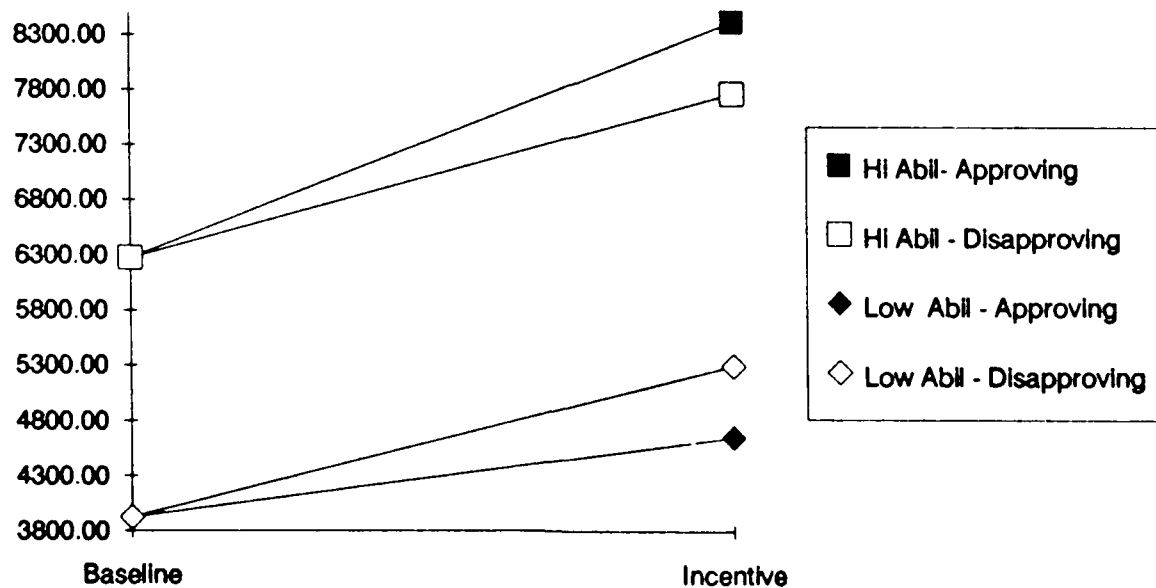


Figure 2. Performance of high and low ability operators in 12 person groups which approve and disapprove of reward system.

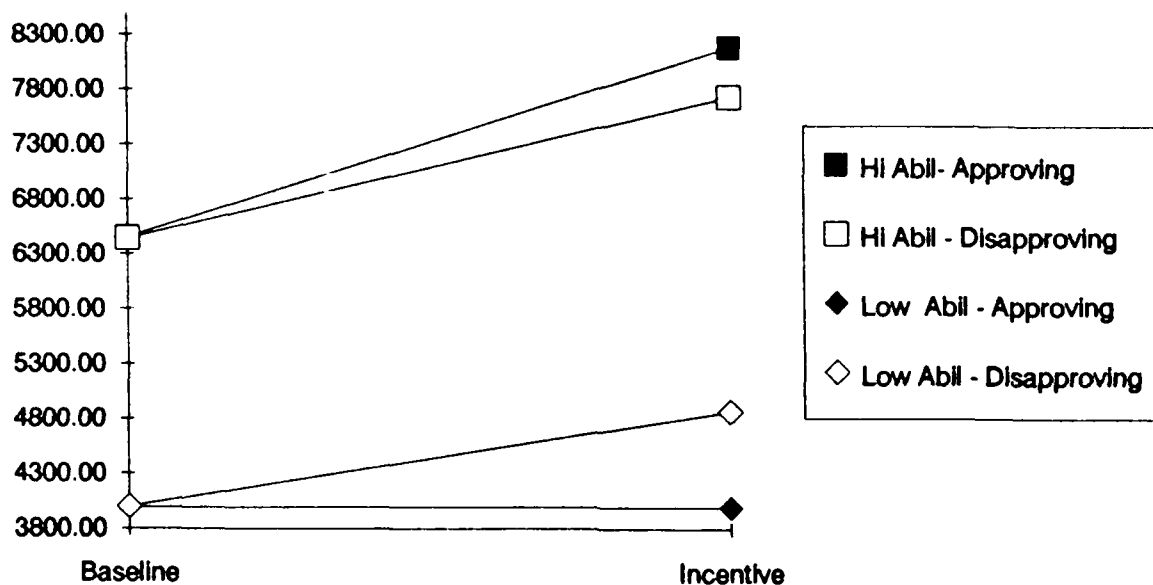


Figure 3. Performance of high and low ability operators in 24 person groups which approve and disapprove of reward system.

The size of this increase depends on the ability of the worker, whether the individual is a member of an approving group, and whether the individual is a member of the six person group. High ability operators were significantly more influenced by the approval of their groups members than the low ability operators. This difference is so strong as to suggest that low ability operators actually do better when their group members disapprove of Gain Sharing in 12 and 24 person groups! In the six person groups, the high ability operators increased performance significantly when the group members approved of Gain Sharing than in the other groups.

Discussion and Conclusions

These results offer help in answering the original questions posed

by this research and development. The effects of Gain Sharing are moderated by group size and the degree of group approval of the Gain Sharing system. In addition worker ability plays an important role in how size and approval influence performance. These results also offer a possible explanation of what previously appeared to be discrepant findings. The social loafing literature (Latane, Williams, & Harkins, 1979) has reported that as group size increases group members will reduce their efforts and consequently their performance. While White's (1979) review of the Gain Sharing literature could not find a similar result. The possible explanation of these discrepant findings may be the result of group approval of the Gain Sharing systems and in the ability of the workers.

It appears that when workers are not very proficient in the group's task, they take advantage of the groups

enthusiasm and take a free ride. This is particularly true when the group is large and the individual is relatively anonymous. Highly proficient workers on the other hand respond favorably to the groups approval and expend extra effort to increase performance. This might be described as a "band wagon" effect. This is especially true in smaller groups where they may be more visible and perhaps recognizable for their achievements.

From a practical perspective these results have several important applications:

1. Gain Sharing systems have the potential to make substantial improvements to Navy productivity; they should be developed and implemented wherever practical.

2. The large effect of worker ability or proficiency on performance, even after the introduction of Gain Sharing, suggests that efforts to increase the proficiency of Navy employees is vital. Attracting, selecting, training, and developing Navy workers should have high priority.

3. Before introducing a Gain Sharing system in an organization, it is important to establish that group members approve of the system and are willing to work to make it succeed. This is particularly true when the group size is small.

4. In larger groups and especially groups where there is high member anonymity, feedback about member approval will only improve the performance of the highly proficient workers. The information may have a detrimental effect on the less proficient workers and give them reason to look for a free ride.

5. Gathering and reporting worker approval information to group members may be justified just because of the effect it has on highly proficient workers. However, an alternative may be worth considering. While the data reported here do not speak directly to the matter, in large groups, it may be preferred to collect and report the degree of approval for a Gain Sharing system and group performance in small subunits of the group as a whole. This is likely to be particularly useful when the subunit has face to face contact or has established working interrelationships. In this way, some of the advantages of smaller group size may be found in the largest of Gain Sharing systems.

The results from this research have been found valuable in our understanding of how to improve the Navy's productivity. A large number of important questions still await research and development. The method of conducting organizational simulations continues to be a cost effective way to answer many of these questions prior to field testing.

References

- Bullock, R. J., & Bullock, P. F. (1982). Gain sharing and Rubik's Cube: Solving system problems. *National Productivity Review*, 1(4), 396-407.
- Boyett, J. H. (1987, Winter). Gain sharing: How to pay for performance. *The Tarkenton Productivity Update*, 6(1), 1-8.
- Davis, J. H., (1969). Group performance. Reading, MA: Addison-Wesley.
- Hammerstone, J. E. (1987). How to make gain sharing pay off. *Training and Development Journal*, 41(4), 80-81.

- Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 21, 822-832.
- Lawler, E. E. (1985). *Gain sharing research: Findings and future directions*. (Rep. No. T 85-1(67)). Los Angeles: University of Southern California, Center for Effective Organizations.
- Marriot, R. (1949). Size of working group and output. *Occupational Psychology*, 23, 47-57.
- Mohr, D. A., Riedel, J. A., & Crawford, K. S. (1985). *A group wage incentive system for production workers at Pearl Harbor Naval Shipyard: Test and evaluation*. (Tech. Rep. No. 85-35). San Diego: Navy Personnel Research and Development Center.
- Nebeker, D. M., Neuberger, B. M., & Hulton, V. N. (1983). *Productivity improvement in a purchasing division: Evaluation of a performance contingent reward system (PCRS)*. (Tech. Rep. No. 83-84). San Diego: Navy Personnel Research and Development Center.
- Ross, T. L., & Hauck, W. C. (1984). Gain sharing in the United States. *Industrial Management*, 26(2), 9-14.
- Schuster, M. (1984). The Scanlon plan: A longitudinal analysis, *Journal of Applied Behavioral Science*, 21(1), 23-38.
- White, J. K. (1979). The Scanlon plan: Causes and correlates of success, *Academy of Management of Journal*, 22(2), 292-312.
- See pages B-3 and B-4 for FY88 publications.

APPENDIX A
PROJECT TRANSITIONS

Transitions

Independent Research

Brain mechanisms for human color vision: Implications for display systems (R000-N0-000-01) applicable to FY88/89 USMC exploratory development 6.2 research.

Stabilization of performance on a computer-based simulation of a complex cognitive task (R000-N0-000-03) will transition to advanced development 6.3 computer-based performance testing.

Results from the project titled *Policy modeling techniques for large-scale multiple objective problems* (RR000-01-042-025) will transition to exploratory development 6.2 work in FY89.

Independent Exploratory Development

Reading comprehension strategies (RV36-I27-02) ideas and strategies explored during this effort will transition to advanced development 6.3 *Prerequisite skills enhancement program*.

Results from the project *Group size and member approval of reward plans in a gain sharing system: Effects on individual performance* (RV36-I27-04) included the development and presentation of OCPM and ASN(S&L) sponsored training course entitled "An Introduction to Productivity Gain Sharing." This course is being offered across the Navy to organizations interested in adopting a Productivity Gain Sharing System for improving their organizational productivity. We are also providing technical assistance to five large Navy organizations as well as help them adopt gain sharing systems under the direction and sponsorship of ASN (S&L).

The sea/shore rotation model formulated under the project *Optimal control theory for system of quasi-linear difference equations* (RV36-I27-01) was refined and developed into a computer algorithm and transitioned to engineering development program element 0604703N as an integral part of the Sea/Shore Rotation Management System under development to support the Navy Enlisted Personnel Rotation System (NEPERS). NEPERS is the Navy's radically new rotation program recently approved for development and testing by the DCNO for Manpower (OP-01).

Loss forecasting with empirical Bayes estimators (RV36-I27-05) will transition into existing project 6.2 Marine Corps force management forecasting and two existing 6.3 projects, Marine Corps enlisted planning system and the Navy's distributable inventory management information system.

APPENDIX B

PRESENTATIONS AND PUBLICATIONS

Presentations

Independent Research

Ali, A. I., Kennington, J. L., Liang, T. T., Thompson, T. J. (1988). *Network models and algorithms for Navy personnel assignment*. Invited presentation at the Joint National Meeting of the Operations Research Society of America and the Institute of Management Sciences, Washington, DC.

Kidder, P. J. (In press). Participation and satisfaction in the performance assessment process. *Proceedings of the 29th Annual Conference of the Military Testing Association*. Ottawa, Ontario, Canada.

Liang, T. T., & Buclatin, B. B. (1988). *An approach to improve personnel unit readiness*. 56th Symposium of the Military Operations Research Society, Monterey, CA.

Independent Exploratory Development

Baker, M. S. (1988). *Reading comprehension strategies*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Haladyna, T. M., & Sympton, J. B. (April 1988). *Empirically-based polychotomous scoring of multiple-choice items: Historical overview*. Paper presented in C. E. Davis (Chair), "New developments in polychotomous item scoring and modeling." Symposium conducted at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Holmes, R. (1988). *Improving least squares estimates via empirical Bayes*. Talk given to the 1988 Annual Meeting of the American Statistical Association, New Orleans, LA.

Nebeker, D. M. (1988). *A comparison of total quality management and goal setting in an organizational simulation*. Paper presented presentation at the August 1988 Annual Convention of the Academy of Management, Anaheim, CA.

Nebeker, D. M. (1988). *Computer monitoring: What are its effects on workstation performance, satisfaction, and stress?* Part of the symposium "Impacts of Electronic Monitoring." Presented at the August 1988 Annual Convention of the Academy of Management, Anaheim, CA.

Sympton, J. B. (June 1988). *Giving credit where credit is due: A new approach to test scoring*. Invited talk presented in the Navy Personnel R&D Center Interdepartmental Seminar Series, San Diego, CA.

IR/IED FY83 Annual Report

Simpson, J. B. (May 1988). *A procedure for linear polychotomous scoring of test items*. Paper presented at the ONR Conference on Model-based Psychological Measurement, Iowa City, IA.

Simpson, J. B. (August 1988). *Polyweighting: A new approach to scoring personnel tests*. Invited talk for members of the Personnel Testing Council of Northern California, Oakland, CA.

Simpson, J. B. (July 1988). *Polyweighting: A new approach to scoring personnel tests*. Invited talk for members of the Personnel Testing Council of Southern California, Los Angeles, CA.

Simpson, J. B., & Haladyna, T. M. (April 1988). *An evaluation of polyweighting in domain-referenced testing*. Paper presented in C. E. Davis (Chair), "New developments in polychotomous item scoring and modeling." Symposium conducted at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Publications

Independent Research

- Ali, A. I., Kennington, J. L., & Liang, T. T. (1988). Assignment with en route training of Navy personnel. *Naval Research Logistics*.
- Chang, F. R. (In press). Writing text for many readers: A role for technology. In D. Wagner (Ed.), *The future of literacy in a changing world: Syntheses from the industrialized and developing nations*. Elmsford, NY: Pergamon Press.
- Federico, P-A. (In press). Student cognitive attributes and performance in a computer-managed instructional setting. In R. F. Dillon and J. W. Pellegrino (Eds.), *Testing, Volume 2: Theoretical and applied perspectives*. New York: Greenwood Press.
- Liang, T., & Buclatin, B. (January 1988). Improving the utilization of training resources through optimal personnel assignment in the U.S. Navy. *The European Journal of Operational Research*, 33(2), 183-190.
- Trejo, L. J., & Lewis, G. W. (In press). Sensitivity to hue differences measured by visual evoked potentials. In *Proceedings of the First Navy Independent Research/Independent Exploratory Development Symposium*. Laurel, MD: Chemical Propulsion Information Agency.

Independent Exploratory Development

- Boyle, J. P. (1988). *An empirical Bayes approach to forecasting Marine Corps enlisted personnel loss dates* (NPRDC Tech. Rep. 88-54). San Diego: Navy Personnel Research and Development Center.
- Holmes, R., & Boyle, J. P. (In press). Improving least squares estimates via empirical Bayes. In the *American Statistical Association 1988 Proceedings of the Business and Economics Statistics Section*. New Orleans, LA.
- Krass, I. (In press). An application of dynamic modeling to the sea shore rotation planning problem in the Navy. *Computers and Mathematics With Applications*.
- Montague, W. E. (1988). Promoting cognitive processing and learning by designing the learning environment. In Jonassen (Ed.), *Instructional designs for microcomputer courseware*. Hillsdale, NJ: L. Erlbaum Associates.
- Nebeker, D. M. (In process). *The effects of different performance-reward functions on performance satisfaction and stress*. San Diego: Navy Personnel Research and Development Center.

Nebeker, D. M. (In process). *Work standards, productivity, and quality*. San Diego: Navy Personnel Research and Development Center.

Riedel, J. A., Nebeker, D. M., & Cooper, B. L. (In press). The influence of monetary incentives on goal choice, goal commitment, and task performance. *Organizational Behavior and Human Decision Processes*.

APPENDIX C

AWARDS AND HONORS

Awards and Honors

Independent Research

Trejo, L. J., & Lewis, G. W. *Sensitivity to hue differences measured by visual evoked potentials* was nominated for the Best Navy Independent Research Paper of 1988 and the authors received a commendation from RADM Wilson, Chief of Naval Research.

DISTRIBUTION LIST

Distribution:

Assistant Secretary of Defense
Deputy Under Secretary of Defense for Research and Engineering (Research and
Advanced Technology)
Assistant for Training and Personnel Technology (OUSD) (A)/R&T(E&LS)
Assistant Secretary of the Navy (Manpower and Reserve Affairs) (OASN) (M&RA)
Deputy Assistant Secretary of the Navy (Manpower)
Assistant for MPT Research and Development and Studies (OP-01B7)
Assistant for Manpower, Personnel, and Training (OP-987H)
Commander, Naval Sea Systems Command
Commander, Naval Air Systems Command
Commander, David Taylor Ship Research and Development Center
Commanding Officer, Naval Coastal Systems Center
Commanding Officer, Naval Underwater Systems Center
Commander, Naval Ocean Systems Center
Commander, Naval Air Development Center
Officer in Charge, Naval Surface Warfare Center Detachment, White Oak
Laboratory
Commander, Naval Weapons Center (Code 38)
Life Sciences Directorate (OCNR-114)
Office of Naval Research, Detachment Pasadena
Office of Naval Research, Detachment London
Chief of Naval Education and Training (Code 00A)
Chief, Bureau of Medicine and Surgery (MED 25)
Commanding Officer, Naval Aerospace Medical Research Laboratory
Commandant of the Marine Corps (MPI-20)
Commandant of the Marine Corps (RD)
Commander, U.S. ARI, Behavioral and Social Sciences (PERI-POT-I)
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base, TX
Program Manager, Life Sciences Directorate, Bolling Air Force Base (AFOSR/NRL)
Commander, Air Force Human Resources Laboratory, Wright-Patterson Air Force
Base (AFHRL/LRS-TDC)
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base
(AFHRL/ID)
Director of Training, Office of Civilian Personnel Management
Superintendent, Naval Postgraduate School
Director of Research, U.S. Naval Academy
Executive Director, U.S. Naval Institute
Institute for Defense Analyses, Science and Technology Division
Center for Naval Analyses, Acquisition Unit
Defense Technical Information Center (DTIC)

Copy to:

Director of Navy Laboratories Office (SPAWAR 05T), (SPAWAR 51E1)
Life Sciences Directorate (OCNR-30T), (OCNR-0722)

END
FILMED
6-89
DTIC